

**On Improving the Acquisition and Reconstruction Of  
Spatio-Temporal Magnetic Resonance Imaging**

by

Siddharth Srinivasan Iyer

S.B., University of California Berkeley (2017)  
S.M., Massachusetts Institute of Technology (2019)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

© Siddharth Srinivasan Iyer, MMXXII. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly  
paper and electronic copies of this thesis document in whole or in part in any medium  
now known or hereafter created.

Author .....  
Department of Electrical Engineering and Computer Science  
August 26, 2022

Certified by .....  
Kawin Setsompop  
Professor of Radiology and Electrical Engineering, Stanford University  
Thesis Supervisor

Accepted by .....  
Leslie A. Kolodziejski  
Professor of Electrical Engineering and Computer Science  
Chair, Department Committee on Graduate Students



# **On Improving the Acquisition and Reconstruction Of Spatio-Temporal Magnetic Resonance Imaging**

by  
Siddharth Srinivasan Iyer

Submitted to the Department of Electrical Engineering and Computer Science  
on August 26, 2022, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## **Abstract**

Magnetic Resonance Imaging (MRI) is a non-invasive but slow imaging modality that provides unparalleled flexibility in acquiring multiple forms of soft-tissue contrast. Recently, there has been a lot of interest in mapping the inherent magnetization properties of the underlying human tissue and in temporally resolving the acquired data. Broadly classified as spatio-temporal MRI, these methods yield unprecedented details of the human anatomy and function, improving clinical diagnostic performance and prognosis. However, such methods are inherently high-dimensional, resulting in encoding-intensive data acquisition processes and computationally-intensive reconstructions. This begets long acquisition and reconstruction times, making such methods difficult to integrate into clinical workflows. This thesis aims to improve the acquisition and reconstruction times of spatio-temporal MRI to enable its use in clinical and neuroscientific setting.

Thesis Supervisor: Kawin Setsompop

Title: Professor of Radiology and Electrical Engineering, Stanford University



## Acknowledgments

I would like to thank my advisor, Kawin Setsompop, for his guidance through the last five years. He is and has always been super supportive, and his passion to always push the boundaries of what is possible has been inspiring. Through my time at Kawin's lab, I have met many fantastic peers I would now call my friends. Congyu Liao is one of the most driven people I have met, with no restraint in reaching his goals. Xiaozhi Cao has always been super helpful, and will always take time out to help me in MRI sequence programming. Sophie Schauman helped me a lot during the final stretch of this thesis, and I am not sure if I could have finished this without her help. Nan Wang has been a fantastic collaborator to work with, and I really enjoy our discussions on the theoretical side of reconstruction. Working with Quan Chen is really fun as we often dive into the deep end of mathematical equations. Ariel Hannum is always on top of getting snacks for the lab, and we perhaps spent a bit too long talking about different ways of making figures. Mahmut Yurt is the resident deep learning wizard, and is always ready to take on challenges head on. Merlin Fair was perhaps the only person I could talk about cricket to. Jaqueline Velazquez has really helped me with the administration side of things. Fuyixue Wang and Zijing Dong have unrelenting work ethics that are inspiring. Daniel Polak helped me get started with MRI sequence programming. Berkin Bilgic helped me feel welcome during my first few years of graduate studies. While we have not worked extensively on a project together, I am thankful to have met Itamar T, Praneeth M, Shirong W, Yannick B, Molin Z, Avery B, Gilad L, Jason S, Mary-Kate M, Sohyun H, Stephen C, Jon P and Thomas W.

I would like to thank my committee members, Elfar Adalsteinsson and Luca Daniel. I had a great experience working with you, and I really enjoyed explaining the nuances to my projects to such an attentive committee.

I would also like to thank my undergraduate advisor, Michael (Miki) Lustig. It was his passion in imaging that inspired me to get into this field to begin with. He also made me feel like I was a valued member of the lab even as an undergraduate. It was in Miki's lab that I met many inspiring people that have greatly influenced my outlooks. Martin Uecker inspired me to think about computation and Mariya Doneva was very patient with me as I struggled through writing my first manuscript. Jonathan (Jon) Tamir and Frank Ong have been fantastic mentors and friends. Jon's determination to push his research into the clinic has been inspiring, and Frank's ability to walk the line of "really cool theory but also here is how it is practical" is something I still aspire to be able to replicate.

I would also like to thank Enhao Gong and Zechen Zhou for a great internship experience at Subtle Medical.

I would like to thank my friends for their help and support throughout the years. In no particular order: Ajin Tom, Rishabh Karajgi, Caroline Chen, Jordan French, Eva Malis, Allie Lalor, Charlie James, Grant Wang, Anurag Ajay, Rohan Chitnis, Yamin Arefeen, Mia Polansky, Luke Van Der Spoel, Suma Anand, Jessica Van Brummelen, Sherdil Niyaz, Charles Zhang, Sudip Guha, Julio Oscanoa, Zhitao Li, Chris Sandino, Zain Chughtai, Elizabeth Cole, Aditi Kulkarni, Aneesha Kulkarni, Preetum Nakkiran, and Jorge Garcia.

I would also like to thank all of my teachers through the years for all their support and guidance.

I would like to dedicate this to my family: to my beautiful wife Lucy, to my lovely puppy Bhageera and her older brother Blu, to my always supportive parents: Srinivasan Narayanan and Richa Aiyer, and to my extended family: Amit Kapoor, Carli Kim, Yolanda Kapoor, Ganesh Srinivasan, Kalyani Ganesh, Shyam Srinivasan, Shreya Ganesh, N. Srinivasan, Danam Srinivasan, Sunil Mehra, Alpa Mehra, Harish Kapoor, Urmil Kapoor, Ishan Mehra, Arshia Mehra, Hadley Banas, Hayden Banas, Gary Tate and Truyen Tate.



# Contents

<b>1</b>	<b>Introduction</b>	<b>9</b>
<b>2</b>	<b>Background</b>	<b>15</b>
2.1	Linear Inverse Problems . . . . .	15
2.2	Iterative Proximal Algorithms . . . . .	16
2.3	MR Physics . . . . .	18
2.3.1	Spin Dynamics . . . . .	18
2.3.2	Fourier Modeling . . . . .	18
2.3.3	Contrast in MRI . . . . .	20
2.3.4	Sources of Error . . . . .	20
2.4	MR Encoding . . . . .	21
2.4.1	Parallel Imaging . . . . .	21
2.4.2	Wave Encoding . . . . .	21
2.4.3	Compressed Sensing . . . . .	22
2.4.4	Spatio-Temporal MRI . . . . .	23
<b>3</b>	<b>Polynomial Preconditioning for Regularized Linear Inverse Problems</b>	<b>25</b>
3.1	Introduction . . . . .	25
3.1.1	Contributions . . . . .	25
3.1.2	Related Works . . . . .	26
3.2	Main Result . . . . .	27
3.3	Theory . . . . .	28
3.3.1	Deriving Polynomial Preconditioning for PGD . . . . .	28
3.3.2	Polynomial Design . . . . .	30
3.3.3	Error Bound . . . . .	31
3.3.4	Implementation Details and Complexity Analysis . . . . .	31
3.3.5	Interpretation and Noise Coloring . . . . .	32
3.4	MRI Experiments . . . . .	33
3.4.1	Parameter Selection . . . . .	33
3.4.2	Cartesian MRI . . . . .	34
3.4.3	Non-Cartesian MRI . . . . .	35
3.4.4	Spatio-Temporal MRI . . . . .	36
3.5	Discussion . . . . .	37
<b>4</b>	<b>Deep Learning Initialized Compressed Sensing</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	Background . . . . .	49
4.3	Methods . . . . .	49
4.3.1	Memory Optimization . . . . .	50
4.3.2	Density Compensation . . . . .	50
4.3.3	Field of View Processing . . . . .	51
4.3.4	Deli-CS . . . . .	52
4.3.5	Basis Balancing . . . . .	53

4.4	Experiments . . . . .	54
4.5	Results . . . . .	56
4.6	Discussion . . . . .	56
<b>5</b>	<b>Better Encoding of Spatio-Temporal MRI using Wave-Shuffling</b>	<b>69</b>
5.1	Introduction . . . . .	69
5.2	Theory . . . . .	70
5.2.1	Wave-Encoding Parameter Optimization . . . . .	70
5.2.2	Wave-PSF Calibration . . . . .	71
5.3	Methods . . . . .	73
5.3.1	Subspace Design . . . . .	73
5.3.2	PSF Analysis and Wave-Encoding Parameter Optimization . . . . .	73
5.3.3	In-Vivo Experiments . . . . .	74
5.4	Results . . . . .	76
5.4.1	Subspace Design . . . . .	76
5.4.2	PSF and Parameter Optimization . . . . .	76
5.4.3	In-Vivo Experiments . . . . .	77
5.5	Discussion . . . . .	79
<b>6</b>	<b>Physics Calibration for Spatio-Temporal MRI</b>	<b>85</b>
6.1	Introduction . . . . .	85
6.2	Methods . . . . .	85
6.3	Experiment . . . . .	86
6.4	Results . . . . .	87
6.5	Discussion . . . . .	87
<b>7</b>	<b>Conclusions</b>	<b>91</b>

# 1: Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive but slow imaging modality that provides unparalleled flexibility in acquiring multiple forms of soft-tissue contrast. Advances in MRI are moving the field towards using data acquisition across multiple spatial-temporal dimensions with high-density receiver coils. These methods yield unprecedented details of the human anatomy and function, improving clinical diagnostic performance and prognosis. However, such methods are inherently high-dimensional, resulting in encoding intensive data acquisition processes and computationally expensive reconstructions. This begets long acquisition and reconstruction times, making such methods difficult to integrate into clinical workflows. With the advent of these imaging applications comes the need to develop techniques for more efficient data encoding and image recovery.

The process of image acquisition and reconstruction, and the challenges mentioned above, can be viewed through the lens of the following regularized linear inverse problem.

$$\operatorname{argmin}_x \underbrace{\frac{1}{2} \|Ax - b\|_2^2}_{\text{Data Consistency}} + \lambda g(x) \quad (1.1)$$

Here,  $A$  is a linear function that models how the underlying image  $x$  propagates through the acquisition process to become the acquired data  $b$ . Encoding optimization can be described as the architecting of the data measuring process so that the resulting  $A$  matrix is better conditioned and more accurately models how to image gets mapped to the acquired data. The consequent minimization of the “Data Consistency” (DC) term in the objective function leverages the better encoding of  $A$  to ensure that the reconstruction more closely describes the acquired data. This DC term is often augmented with a regularization function  $g$  that enforces priors on the image during reconstruction, where the regularization value  $\lambda$  controls the strength of the prior with respect to the DC term. This is done to aid reconstruction when the acquired data is corrupted by noise and/or the  $A$  matrix is not injective or ill-posed.

The large acquisition matrix  $A$  of MRI makes gradient-based algorithms the go-to methods for solving (1.1). In particular, iterative proximal gradient techniques have seen wide spread adoption due to their ease-of-use and theoretically optimal properties. However, these algorithms are adversely affected by the distribution of eigenvalues of  $A^*A$ , which translates to slow iterative convergence and long reconstruction times. This is particularly true for spatio-temporal MRI applications.

Note that in the above context and for the rest of the thesis, “time-resolved” is not related to real-time acquisitions, but rather the recovery of the magnetization temporal dynamics of the underlying human tissue.

## Outline

This thesis aims to enable the robust clinical translation of spatio-temporal MRI by jointly considering both the acquisition and algorithmic reconstruction for overall faster acquisition and reconstruction times.

*Chapter 2* provides a brief background of the topics covered in this thesis, and is split into four sections that attempt to partition the topics covered in this thesis for ease of reference. In particular, the first paragraph of the following chapters lists which of the four sections are “prerequisites” (used very loosely) and “recommended readings” to hopefully aid the readability of this thesis.

*Chapter 3* studies the iterative proximal gradient methods typically used to solve (1.1), and proposes

a generalizable polynomial-based preconditioner to accelerate iterative convergence, resulting in faster reconstruction times. While focused in MRI reconstruction, this method is highly general and can be applied to various imaging applications of interest.

*Chapter 4* further reduces reconstruction times by using deep-learning to jump-start the above iterative algorithm. This initialization approach is synergistic in that the iterative algorithm guards against deep learning hallucinations while still leveraging the fast processing times of machine learning to reduce the required number of iterations of the reconstructions. This chapter also discusses practical challenges that come with deploying spatio-temporal MRI methods to the clinic.

*Chapter 5* switches gears into the optimizing the MRI encoding. This chapter studies a specific spatio-temporal MRI method that reconstructs a time-series of images at multiple clinical contrasts, and augments it with an improved encoding technique that better utilizes the MRI receive channel information to reduce the acquisition time.

Lastly, *Chapter 6* presents an ultra-fast unified, rapid calibration acquisition process termed Physics Calibration (PhysiCal). This uses carefully designed sampling across the temporal dimension for robust and accurate recovery of physics-based parameter maps, which enable the better modeling of  $A$  in (1.1) for improved accuracy of spatio-temporal MRI.

## Contributions

### Conference Proceedings

- S. Iyer, C. Sandino, M. Yurt, X. Cao, C. Liao, S. Schauman and K. Setsompop. “SMILR - Subspace Machine Learning Reconstruction” In Proceedings of the International Society of Magnetic Resonance in Medicine, 2022. Invited Talk.
- S. Iyer, F. Ong and K. Setsompop. “Polynomial Preconditioning for Accelerated Convergence of Proximal Algorithms including FISTA” In Proceedings of the International Society of Magnetic Resonance in Medicine, 2022.
- S. Iyer, C. Liao, Q. Li, M. Manhard, A. Berman, B. Bilgic and K. Setsompop. “PhysiCal: A rapid calibration scan for  $B_0$ ,  $B_1^+$ , coil sensitivity and Eddy current mapping.” In Proceedings of the International Society of Magnetic Resonance in Medicine, 2020. Invited Talk.
- S. Iyer, D. Polak, C. Liao, S. Cauley, B. Bilgic and K. Setsompop. “Rapid, Time-Resolved Brain Imaging with Multiple Clinical Contrasts using Wave-Shuffling.” In Proceedings of the International Society of Magnetic Resonance in Medicine, 2019. Invited Talk.
- S. Iyer, B. Bilgic and K. Setsompop. “Faster ( $T_2$ -)Shuffling With Wave-Encoding.” In Proceedings of the International Society of Magnetic Resonance in Medicine, 2018. Invited talk.

### Manuscripts

- S. Iyer, D. Polak, C. Liao, J. Tamir, S. Cauley, B. Gagoski, W. Lo, B. Bilgic and K. Setsompop. “Wave-encoding and Shuffling Enables Rapid Time Resolved Structural Imaging.” Submitted to Magnetic Resonance in Medicine. Preprint: <https://arxiv.org/abs/2103.15881>
- S. Iyer, F. Ong, X. Cao, C. Liao, L. Daniel, J. Tamir and K. Setsompop. “Polynomial Preconditioners for Regularized Linear Inverse Problems.” Submitted to Society of Industrial and Applied Mathematics (SIAM) Journal of Imaging. Preprint: <https://arxiv.org/abs/2204.10252>
- S. Iyer, S. Chauman, C. Sandino, M. Yurt, X. Cao, C. Liao and K. Setsompop. “Deep Learning Initialized Compressed Sensing (Deli-CS) with application in Rapid Volumetric Spatio-Temporal Subspace Reconstruction.” In preparation.
- S. Iyer, X. Cao, C. Liao, Q. Li, M. Manhard, A. Berman, B. Bilgic and K. Setsompop. “PhysiCal - Rapid Calibration Scan for  $B_0$ ,  $B_1^+$  and Coil Sensitivity Information.” In preparation.

## Assumptions and Conventions

Unless otherwise stated, the following conventions and assumptions hold true for all the following discussions.

- Complex numbers are always the underlying field and vector space.
- The assumed vector spaces are always finite dimensional.
- $\langle \cdot, \cdot \rangle$  denotes standard Euclidean complex inner product.
- Given a linear function  $A$ , its adjoint is denoted as  $A^*$ .
- The normal operator of a linear function  $A$  is denoted  $A^*A$ .
- $\|\cdot\|_p$  denotes the standard Euclidean  $l_p$ -norm.
- $\|\cdot\|_{p \rightarrow p}$  denotes the induced  $l_p$ -norm.
- Unless otherwise stated, the  $l_2$ -induced norm of a linear function  $A$  is unitary. That is to say,

$$\|A\|_{2 \rightarrow 2} = 1.$$

## Abbreviations

- MR Magnetic Resonance.
- MRI MR Imaging.
- CG Conjugate Gradient.
- GD Gradient Descent.
- prox Proximal operator [76].
- PGD Proximal Gradient Descent [76].
- ISTA The Iterative Shrinkage-Thresholding Algorithm [76].
- FISTA Fast ISTA [4].
- PDHG Primal-Dual Hybrid Gradient [19].
- ADMM Alternating Direction Method of Multipliers [76].
- IFISTA Improved FISTA [113].
- DC Data Consistency.
- LLR Locally Low-Rank [94].
- SENSE Sensitivity Encoding for Fast MRI [85].
- ESPIRiT Eigenvalue Self-Consistent Parallel Imaging Reconstruction [96].
- JSENSE Joint Image Reconstruction and Sensitivity Estimation in SENSE [109].
- CPU Central Processing Unit.
- GPU Graphics Processing Unit.
- FSE Fast Spin Echo [67].
- MPRAGE Magnetization-Prepared Rapid Gradient-Echo [68].
- NRMSE Normalized Root Mean Squared Error.
- TR Repetition time.
- TE Echo time.
- TF Turbo factor.
- PSF Point Spread Function.
- SVD Singular Value Decomposition.
- BW Bandwidth.
- TI Inversion time.
- ESP Echo spacing.
- SNR Signal to noise ratio.
- GRE Gradient echo.
- EPI Echo planar imaging.

GLM Generalized linear modelling.

BS Bloch-Siegart.

ACS Auto-calibrated signal.

ETL Echo train length.

FOV Field of View.

NUFFT Non-uniform Fourier Transform.

## 2: Background

This chapter provides a succinct introduction to the topics considered in this thesis. First, the framework of linear inverse problems is discussed along with typical optimization methods used to solve the same. Next, a brief overview of MR acquisition is presented, followed by an overview of MR reconstruction as a linear inverse problems.

### 2.1: Linear Inverse Problems

Computational sensing applications, particularly in imaging, perform reconstruction by solving an optimization problem that enforces the result to be consistent with the acquisition model and the measured information [21, 36]. When the measurement process is linear, the formulated optimization problems are commonly denoted “Linear Inverse Problems”. If  $x$  is the true underlying signal and  $A$  is the linear acquisition process, the acquired data  $b$  is modelled as follows:

$$Ax = b \tag{2.1}$$

The linear function  $A$  that models the acquisition process is often referred to as either the “forward model” of the application, or the “measurement matrix”. If  $A$  is injective, the image can be reconstructed using simple linear inversion. However, in practice, reconstructing this is not desirable as there are numerous factors that, if not accounted for, may yield sub-optimal results. Some examples include:

1. The linear measurement function  $A$  is not injective, resulting in many possible solutions to (2.1).
2. The measuring process is noisy, which is reflected in measurement vector  $b$ .
3. The linear function  $A$  is ill-conditioned, in that the ratio of its largest and smallest non-zero singular value is high, resulting in a linear inversion process that is highly susceptible to numerical perturbations.

The process of matrix inversion is also typically expensive for large problems, as is the case in MRI.

In practice, the reconstruction problem is posed as an ordinary least squares problem and is solved using iterative first-order algorithms such as Gradient Descent (GD) and Conjugate Gradient (CG) [47, 98, 91].

$$\operatorname{argmin}_x \frac{1}{2} \|Ax - b\|_2^2 \tag{2.2}$$

If the  $A$  matrix is not injective, there are many solutions to (2.2), and it is common practice to use the reconstruction with the minimum  $l_2$ -norm as the result. This can be calculated by either using the pseudo-inverse of  $A^*A$ , or initializing iterative methods like CG and GD with the zero vector [43].

That being said, it is often not the case that solving (2.2) with the minimum  $l_2$ -norm solution is the optimal reconstruction. Many applications, particularly in sensing and imaging, have found success is enforcing alternative priors to perform better reconstruction. These problems, broadly classified as regularized linear inverse problems, solve the following program instead to perform reconstruction.

$$\operatorname{argmin}_x \underbrace{\frac{1}{2} \|Ax - b\|_2^2}_{\text{Data Consistency}} + \lambda g(x) \tag{1.1 revisited}$$

The least squares segment in (1.1) is colloquially referred to as “Data Consistency” (or DC) as it is the component of the objective that enforces the reconstruction to be consistent with the measured data  $b$ .  $g$  is a convex regularization function that enforces a prior that is dependent on the application. Common examples include the  $l_2$ -norm, where  $g(x) = \|x\|_2^2$ , which helps reduce the sensitivity to numerical perturbations when solving (2.2), and the  $l_1$ -norm, where  $g(x) = \|x\|_1$ , which helps enforce sparsity priors on the reconstruction.

Unless otherwise stated, this work assumes without loss of generality that,

$$\|A\|_{2 \rightarrow 2} = 1 \tag{2.3}$$

This is because (1.1) can always be divided by the maximum eigenvalue of  $A^*A$  (estimated using Power Iteration) without affecting the reconstructed result.

## 2.2: Iterative Proximal Algorithms

This section will discuss how regularized inverse problems as in (1.1) are solved using a class of optimization algorithms called iterative proximal methods with [76] used as the main reference. These algorithms have seen wide spread adoption in various imaging and applied mathematics fields due to their general applicability and high level of abstraction, particularly for large dimensional applications where matrix-free<sup>1</sup> implementations of the linear function  $A$  are leveraged.

Given a closed, proper convex function  $g$ , its proximal operator is defined as follows:

**Definition 2.1: Proximal Operator**

$$\mathbf{prox}_{\alpha g}(v) = \left\{ \underset{x}{\operatorname{argmin}} \quad \frac{1}{2} \|x - v\|_2^2 + \alpha g(x) \right. \tag{2.4}$$

This proximal operator can be used to solve optimization algorithms of the form (1.1), with arguably the simplest algorithm being Proximal Gradient Descent (PGD) [76]. With the assumption that the induced norm of  $A$  is unitary, the PGD algorithm to solve (1.1) is listed as Algorithm 2.1.

**Algorithm 2.1: Proximal Gradient Descent [76]**

*Inputs:*

→ forward model $A$	→ Proximal operator: $\mathbf{prox}_{\lambda g}$
→ measurements $b$	→ $x_0 = 0$

*Step  $k$ : ( $k \geq 0$ ) Compute*

$$x_{k+1} = \mathbf{prox}_{\lambda g}(x_k - A^*(Ax_k - b)) \tag{2.5}$$

Perhaps the most prolific example of using proximal operator to solve (1.1) is when  $g$  is the non-smooth  $l_1$ -norm, in which case its proximal operator is the *Soft-Thresholding* operator. In fact, PGD with the soft-thresholding operator is so ubiquitous that it is often referred to instead “ISTA”, or the iterative soft-thresholding algorithm. Please see [76, 4] and references therein for more information.

Other than the simplicity of implementing PGD (Algorithm 2.1), one of its main benefits is that it is relatively easy to test different possibly non-smooth regularization functions  $g$  as long as its proximal operator can be evaluated. PGD will always leverage the smoothness of the DC cost to at least achieve sub-linear global rate of iterative convergence [4].

Before proceeding, it is worthwhile to briefly mention some properties of proximal operators that will be leveraged in the sequel. With respect to the optimality, the stationary point of the iterations (2.5) is a solution to (1.1)[76]. This is stated as Theorem 2.1.

<sup>1</sup>Matrix-free implies the individual entries of  $A$ , or the coefficient matrix, are not stored explicitly in memory thus saving significant computational resources. Instead, the matrix-vector products of  $A$  and  $A^*$  are calculated on the fly.

### Theorem 2.1: Fixed Point Iteration of PGD [76]

A solution  $x^*$  is the optimal solution to (1.1) if and only if

$$x^* = \mathbf{prox}_{\lambda g}(x^* - A^*(Ax^* - b)) \quad (2.6)$$

Another useful property of proximal operators that will be utilized in the sequel is the *firm non-expansiveness* property, listed as Lemma 2.1.

### Lemma 2.1: Firm Non-Expansiveness Property [76]

For all  $x, y \in \mathbb{C}^n$ ,

$$\|\mathbf{prox}_{\lambda g}(x) - \mathbf{prox}_{\lambda g}(y)\|_2 \leq \|x - y\|_2 \quad (2.7)$$

Lemma 2.1 can be used to derive a simple proof that PGD will eventually converge.

### Theorem 2.2: Convergence of Algorithm 2.1

When solving (1.1), assuming a solution exists, Algorithm 2.1 will always converge.

*Proof.* Let  $x^*$  be the solution and  $e_k = x_k - x^*$  be the error over iterations. Subtracting (2.5) with (2.6), taking the  $l_2$ -norm on both sides and utilizing Lemma 2.1 yields:

$$\|e_{k+1}\|_2 \leq \|(I - A^*A)e_k\|_2 \quad (2.8)$$

The sequence  $\{\|e_k\|_2\}$  is a non-increasing sequence that is bounded below, thus the monotone convergence theorem applies.

A natural extension of Theorem 2.2 is that upper bound on the degree of improvement between iterations has a natural interpretation with respect to the spectrum of  $A^*A$ .

### Lemma 2.2: Spectral Interpretation of PGD

Consider  $e_k$  as in (2.8). Splitting  $e_k$  into  $s_k + t_k$  where  $s_k \in \text{null}(A)$  and  $t_k \in \text{null}(A)^\perp$  results in:

$$\|e_{k+1}\|_2^2 \leq \|(I - A^*A)t_k\|_2^2 + \|s_k\|_2^2 \quad (2.9)$$

Therefore, the decrease in error of  $t_k$  is upper-bounded by the eigenvalue spectrum of  $I - A^*A$ . This bound is strict when  $A^*A$  is injective.

If  $A^*A$  is badly conditioned in that the smallest non-zero eigenvalue is significantly smaller than one, the component of the error vector  $e_k$  (as defined in the proof of Theorem 2.2) that corresponds to that small eigenvalue suffers from slow convergence.

This thesis focuses the Nesterov accelerated variant of PGD called FISTA [4, 70] as it enjoys theoretically optimal iterative convergence and thus generalizes well to most cases. That being said, no one algorithm is best suited for every imaging problem, and there are numerous algorithms to look into that may leverage application specific properties to be more performant. Examples of algorithms to look into are ADMM [76], PDHG [19] and many more. In applications where the regularization ( $g$ ) is cognate to the  $l_1$ -norm to enforce sparsity, Iteratively Re-Weighted Least Squares (IRWLS) [36, 29, 107, 20] are also well worth consideration.

To improve the iterative convergence over PGD, FISTA [4, 70] smartly chooses different point  $z_k$  (instead of  $x_k$ ) to evaluate (2.5). Listed as Algorithm 2.2 is FISTA (with  $\beta_k$  proposed in [18]) to solve (1.1).

**Algorithm 2.2: FISTA[4] ( $\beta_k$  from [18])***Inputs:*

- forward model  $A$
- measurements  $b$

- Proximal operator:  $\mathbf{prox}_{\lambda g}$
- $x_{-1} = x_0 = 0$

*Step  $k$ : ( $k \geq 0$ ) Compute*

$$x_{k+1} = \mathbf{prox}_{\lambda g}(z_k - A^*(Az_k - b)) \quad (2.10a)$$

$$\beta_k = k/(k+3) \quad (2.10b)$$

$$z_{k+1} = x_k + \beta_k(x_k - x_{k-1}) \quad (2.10c)$$

**2.3: MR Physics**

This section will provide a brief overview of the physics used in MRI, using [71] as a foundation. The Fourier acquisition of MRI and the temporal dynamics of the underlying signal are discussed.

**2.3.1 Spin Dynamics**

Atoms with an odd number of protons and/or neutrons possess non-zero spin angular momentum, and consequently have a non-zero magnetic dipole moment. This property is present in Hydrogen atoms, which constitute a large percentage of the human body's composition in the form of water molecules. In the presence of a strong directed external magnetic field, denoted  $B_0$ , most of these magnetic moments align and produce a net magnetic field in the same direction as  $B_0$ , with more moments aligning with stronger fields. By convention, the direction of  $B_0$  is the  $z$ -axis and is called the longitudinal direction, while the perpendicular  $xy$ -axes are called the transverse plane. The net magnetization vector,  $M$ , also experiences precession at frequency  $\omega_0 = \gamma B_0$ , where  $\gamma$  is the Larmor frequency and is an intrinsic property of an atom species. This is depicted in Figure (2-1)(A), where  $M_0$  represents the magnitude of net magnetization vector.

This magnetization  $M$  can be manipulated through the use of radio frequency (RF). RF pulses played at frequency  $\omega_0$  along the transverse plane ( $x, y$ ) excite or "tip" the magnetization  $M$  away from the  $z$ -axis, and the strength and duration of the RF pulse determines the angle of rotation. Figure (2-1)(B) depicts a  $90^\circ$  rotation, where an RF pulse is played to tip  $M$  onto the transverse plane.

The magnetization vector  $M$  can be considered in terms of two components: the transverse vector  $M_{xy}$  on the  $xy$ -plane, and the longitudinal vector  $M_z$ .  $M_{xy}$  continues to precess around the  $z$ -axis, and the changing magnetic field induces an electric current in receive coils due to Faraday's Law. This is the source of the MR signal.

Image contrast in MRI comes from the rate at which  $M$  returns to equilibrium, where  $M$  is aligned parallel to  $B_0$  along the  $z$ -axis. The rate at which  $M_{xy}$  returns to 0 and  $M_z$  returns to  $M_0$  are parameterized by the so-called  $T_2$  and  $T_1$  values respectively, where

$$M_{xy} = M_0 e^{-t/T_2} \text{ and } M_z = M_0 (1 - e^{-t/T_1}) \quad (2.11)$$

This is illustrated in Figure (2-1)(C). The intrinsic temporal dynamics allows MR to target particular contrasts based on ( $T_1, T_2$ ) signal recovery, which differs between tissue types, through scan timing and RF pulse design.

**2.3.2 Fourier Modeling**

While the rotating transverse magnetization  $M_{xy}$  is able to produce a detectable signal, additional spatial encoding is needed to obtain an image. This is achieved by differentiating signal from different spatial

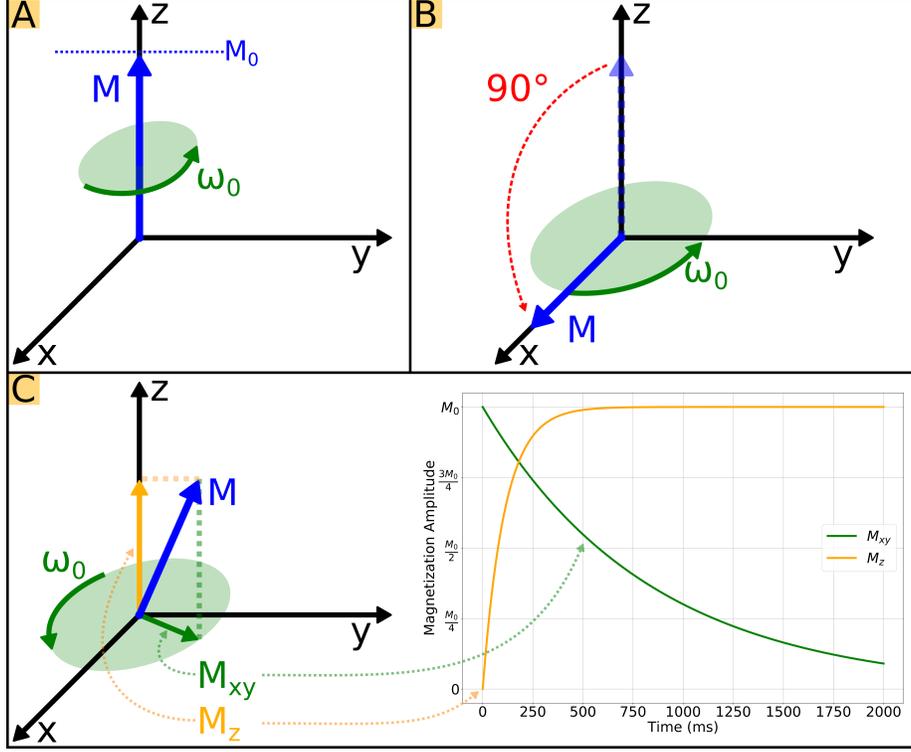


Figure 2-1: (A) In MRI, the magnetization vector  $M$  of Hydrogen atoms in the body align with the strong external  $B_0$  field along the  $z$ -direction.  $M$ , of magnitude  $M_0$ , experiences precession at frequency  $\omega_0 = \gamma B_0$ , where  $\gamma$  is the Larmor frequency and is an intrinsic property of atoms. (B) An RF pulse played at  $\omega_0$  along the transverse plane ( $x, y$ ) with a particular amplitude and duration will tip or rotate  $M$  onto the transverse plane ( $x, y$ ). (B) depicts a  $90^\circ$  rotation. Note that  $M$  will still precess at rate  $\omega_0$ . (C) Over time,  $M$  slowly returns to the thermal equilibrium depicted in (A). The transverse component,  $M_{xy}$  decays over time according to parameter  $T_2$ .  $M_z$  recovers over time according to parameter  $T_1$ . The time recovery curves are described in (2.11).

locations through the use of linear gradient fields that vary the net magnetization in the  $z$ -direction as a function of spatial position ( $x, y, z$ ). Let  $B(t, x, y, z)$  denote the net magnetization in the  $z$ -direction at spatial position ( $x, y, z$ ) at time  $t$ .

$$B(t, x, y, z) = B_0 + G_x(t)x + G_y(t)y + G_z(t)z \quad (2.12)$$

Here,  $(G_x(t), G_y(t), G_z(t))$  are the spatially varying linear gradient fields in the ( $x, y, z$ ) directions respectively.  $B$  prompts the magnetization vector at ( $x, y, z$ ) to precess with frequency:

$$\omega(t, x, y, z) = \gamma(B_0 + G_x(t)x + G_y(t)y + G_z(t)z) \quad (2.13)$$

By changing the frame of reference to a rotating frame at frequency  $\omega_0 = \gamma B_0$ , the  $\omega_0$  term in (2.13) is dropped to simplify the following equations.

Consider the acquisition window from time  $t = 0$  to time  $t = \tau$ , where  $\tau$  is small relative to the relaxation times implying no significant change to  $M$ . The phase, denoted  $\phi$ , accumulated by the precessing magnetization in this window is,

$$\phi_{(x,y,z)}(\tau) = \gamma \left( x \int_0^\tau G_x(t) dt + y \int_0^\tau G_y(t) dt + z \int_0^\tau G_z(t) dt \right) \quad (2.14)$$

Instead of the above, it is common to define  $(k_x(t), k_y(t), k_z(t))$  as follows to simplify the equation.

$$\begin{aligned}\phi_{(x,y,z)}(\tau) &= 2\pi(k_x(\tau)x + k_y(\tau)y + k_z(\tau)z) \\ k_x(\tau) &= \frac{\gamma}{2\pi} \int_0^\tau G_x(t) dt \\ k_y(\tau) &= \frac{\gamma}{2\pi} \int_0^\tau G_y(t) dt \\ k_z(\tau) &= \frac{\gamma}{2\pi} \int_0^\tau G_z(t) dt\end{aligned}\tag{2.15}$$

Let  $m(x, y, z)$  represents the net magnetization at location  $(x, y, z)$ . The detected signal is the sum of the magnetization vectors over all  $(x, y, z)$  locations at time  $\tau$ . Let the transverse magnetization  $M_{xy}$  be encoded onto the complex plane, where the  $x$ -axis and  $y$ -axis form the real and imaginary axis respectively. The acquired signal can then be modeled as:

$$\mathcal{M}(k_x(\tau), k_y(\tau), k_z(\tau)) = \sum_{x,y,z} m(x, y, z) e^{i2\pi(k_x(\tau)x + k_y(\tau)y + k_z(\tau)z)}\tag{2.16}$$

Thus, the signal received at time  $\tau$  represents the spatial frequency of the desired image at position  $(k_x(\tau), k_y(\tau), k_z(\tau))$ . In other words, the spatial frequency representation of the underlying magnetization image ( $m$ ), called  $k$ -space, is the domain which MRI signal is acquired. Let  $\mathcal{F}$  represents the forward Fourier transform. Then,

$$\mathcal{M} = \mathcal{F}m\tag{2.17}$$

Given that  $\tau$  in (2.14) is smaller than typical relaxation times, it is common practice to acquire multiple points within a time-window, denoted “readout”, for faster scan times. If the specified  $(k_x, k_y, k_z)$  are programmed to lie on a Cartesian grid, the  $k_x$  direction is typically called the readout direction with one  $k_x$  line acquired during a readout.

### 2.3.3 Contrast in MRI

Multi-echo MRI acquisitions such as MPRAGE [68] and  $T_2$ -Space [67] leverage specially designed RF pulses to combat  $(T_1, T_2)$  relaxation effects and sustain the signal over a longer period of time, after which a new sequence of RF pulses will need to be played. This reduces the overall scan time. Repetition Time (TR) is the time window within which it is reasonable to acquire signal before a new series of RF pulses is required to re-excite the signal. Within a TR, multiple  $k_x$  lines are acquired, where the center of each  $k_x$  lines is commonly referred to as the “echo time”. Echo Spacing (ESP) is the time between each  $k_x$  line, and the number of echoes acquired in a TR is typically called the Echo Train Length (ETL) or Turbo Factor (TF).

These timing parameters, along with the RF pulse train, largely determine the contrast of the desired image. For example, in MPRAGE, the RF train starts with a  $180^\circ$  magnetization flip followed by a series of small angle rotations, where a  $k_x$  line is acquired after each flip. The Inversion Time (TI) is the time between the first  $180^\circ$  flip and the middle echo (Echo number:  $\lfloor \text{ETL}/2 \rfloor$ ). Along with a small TR, MPRAGE results in images that are dominated by  $T_1$  differences within tissues. In  $T_2$ -Space, the RF train is optimized to sustain signal evolution over a long ETL [67] to acquire images that are dominated by  $T_2$  differences between tissues. This sequence is typically described with respect to its TR and echo time, or TE. TE is the time between the first RF pulse to middle of the first echo, after which adjacent echoes are spaced apart according to ESP.

When leveraging multi-echo images with Fourier encoding as described in the previous sections, the temporal dynamics of the spins are not explicitly considered. This can cause undesirable blurring due to  $k$ -space apodization arising from the  $(T_1, T_2)$  relaxation effects over the echo train.

### 2.3.4 Sources of Error

This section will *very* briefly describe a *small* subset of possible error sources. Fourier encoding is inherently slow (in the order of minutes) making MRI susceptible to motion related artifacts. When scanning a subject, the main  $B_0$  field is not perfectly homogeneous, which can cause inaccuracies when resolving the image using Fourier model, such as signal dropout. Similarly, the applied RF pulse maybe spatially varying, resulting in inaccuracies when estimating  $T_1$  values.

## 2.4: MR Encoding

This section will use the linear inverse problem framework to describe how MRI acquisition is modeled by the composition of linear operators.

### 2.4.1 Parallel Imaging

Real world MRI leverages several receive channels to acquire multiple views of the same underlying data. These channels are arranged in spatially different locations around the anatomy of interest, and thus have different “sensitivities” to the underlying signal. This performs a kind of spatial encoding in addition to the data-redundancy introduced from using multiple receive channels. These properties can be used by the reconstruction for either better noise performance or faster scan times.

In particular, the spatial encoding from the multiple channels enables robust image reconstruction even when the number of k-space points acquired does not satisfy Nyquist criterion. This allows for faster scan times as less data are acquired. Broadly described as Parallel Imaging, the two most common techniques are SENSE [85] and GRAPPA [38], with many more optimized extensions such as [109, 95, 96, 92, 40]. This work will use the Parallel Imaging model described in [85, 96] as it easily integrates into the linear inverse problem framework discussed previously.

Let  $S_1, \dots, S_{n_c}$  denote the spatial sensitivities of the  $n_c$  receive channels,  $y_1, \dots, y_{n_c}$  be the raw k-space data acquired from the respective channels, and  $m$  be the underlying image to be recovered. The acquisition model can be described as follows.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_c} \end{bmatrix} = \begin{bmatrix} \mathcal{M} & 0 & \dots & 0 \\ 0 & \mathcal{M} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathcal{M} \end{bmatrix} \begin{bmatrix} \mathcal{F} & 0 & \dots & 0 \\ 0 & \mathcal{F} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathcal{F} \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_{n_c} \end{bmatrix} m \quad (2.18)$$

$\mathcal{M}$  represents the  $(k_y, k_z)$  sampling mask and  $\mathcal{F}$  is the forward Fourier transform. Each  $S_i$  is arranged as a diagonal matrix. (2.18) is often compactly written as,

$$y = \underbrace{\mathcal{M} \mathcal{F} S}_{A} m \quad (2.19)$$

In this way, MR acquisition is well modelled by (2.1), with  $A$  being the linear acquisition matrix or forward model that maps the underlying signal  $m$  to the acquired data  $y$ .

The degree of under-sampling that can be performed is dependent on how “orthogonal” the spatial sensitivities are to each other. In other words, the encoding-ability of  $S$  is what determines the degree of scan acceleration. When the problem is “well-posed”, in the sense that enough k-space points are acquired given the encoding-ability of the coil, the image  $m$  can be reconstructed using (2.2), typically with CG.

### 2.4.2 Wave Encoding

In achieve further reduction in scan time, Wave-CAIPI [7] utilized additional sinusoidal ( $G_y, G_z$ ) gradients played during the acquisition of a readout (or  $k_x$  line), as depicted in Figure (2-2).

The sinusoidal gradients results in controlled spreading in the  $x$ -direction as a function of spatial position  $(y, z)$ , which in turn results in better utilization of the spatial sensitivity maps  $\{S_i\}$ . The controlled spreading is depicted in Figure (2-2)(B), which shows the effect of the ( $G_y$ ) sinusoidal gradient on the image in the  $(x, y)$  plane. Here, the image voxels are spread along the  $x$ -direction, where the level of spreading increases linearly as a function of absolute position  $(y)$ . This spreading helps increase the spatial distance between aliased voxels in an accelerated acquisition (with under-sampling along  $k_y$ ), which improves the conditioning of the parallel imaging reconstruction (2.19).

With respect to the forward model, the effect of the sinusoidal gradients is implemented with two

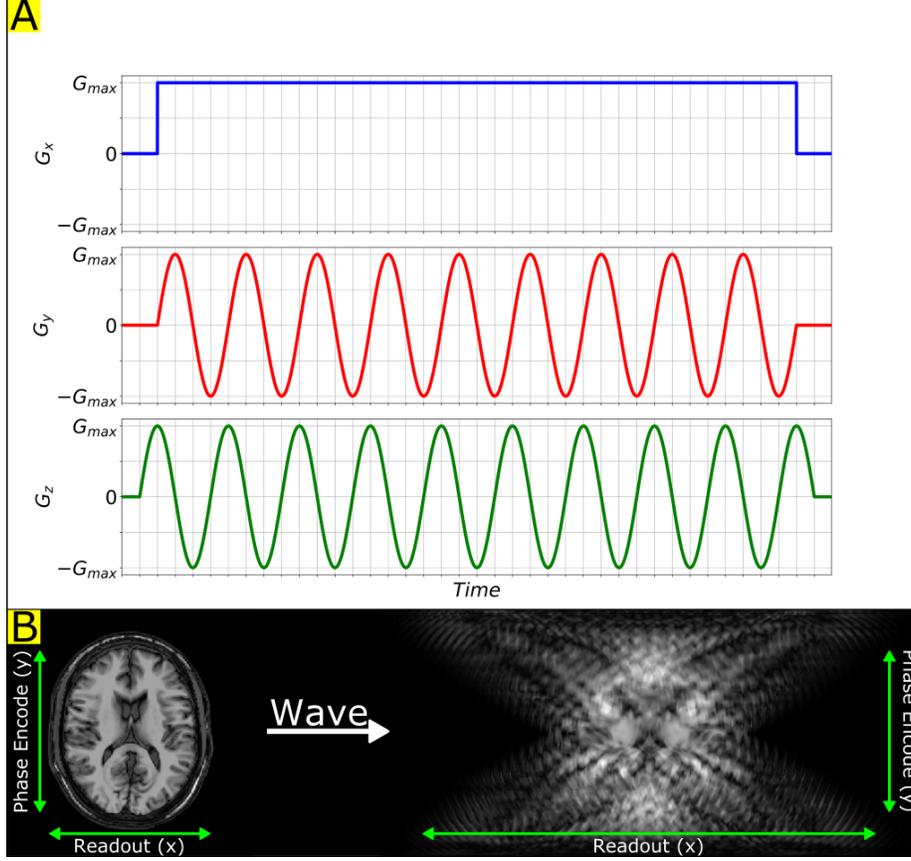


Figure 2-2: (A) depicts Wave’s sinusoidal ( $G_y, G_z$ ) gradients during readout. (B) depicts the controlled spreading in the  $x$  direction as a function of  $(y, z)$  induced by the sinusoidal gradients.

additional operators,  $R$  and  $W$ , that are integrated into (2.19).

$$y = \mathcal{M} \mathcal{F}_{yz} W \mathcal{F}_x R S m \quad (2.20)$$

$\mathcal{F}_x(\cdot)$  is the forward Fourier transform along the  $x$  direction and  $\mathcal{F}_{yz}(\cdot)$  is the forward Fourier transform along the  $(y, z)$  directions.  $R(\cdot)$  is a resize operator that zero-pads the input image along the  $x$  direction, and  $W$  models the  $x$ -spreading as a function of  $(y, z)$ .  $W$  is an element wise multiplication in  $(k_x, y, z)$ , and is constructed as follows.

$$\begin{aligned} W(k_x, y, z) &= e^{-i2\pi(P_y(k_x)y + P_z(k_x)z)} \\ P_y(k_x(t)) &= \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau \\ P_z(k_x(t)) &= \frac{\gamma}{2\pi} \int_0^t G_z(\tau) d\tau \end{aligned} \quad (2.21)$$

This inclusion of sinusoidal gradients non-trivially improves the encoding of the acquisition, allowing for significant reductions in scan times. For example, [82] achieved  $9\times$  faster MPRAGE acquisition on a standard  $3T$  MRI scanner.

### 2.4.3 Compressed Sensing

Compressed sensing in MRI, introduced in [59, 9, 60], exploits the natural sparsity of images through random sampling which in incoherent noise-like artifacts that can be “de-noised” through the use of regularization that promote sparsity in a transform domain. In the case of (2.19), the reconstruction problem

can be formulated as,

$$\begin{aligned} & \underset{m}{\operatorname{argmin}} \quad \|\Psi m\|_1 \\ & \text{such that} \quad y = \mathcal{M}\mathcal{F}Sm \end{aligned} \quad (2.22)$$

$\Psi$  is a sparsity promoting transform, and is chosen depending on the underlying application of interest. The above problem is often reformulated as the following unconstrained problem to leverage the smoothness of the DC term.

$$\underset{m}{\operatorname{argmin}} \frac{1}{2} \|\mathcal{M}\mathcal{F}Sm - y\|_2^2 + \|\Psi m\|_1 \quad (2.23)$$

This formulation falls in the regularized linear inverse problem category in (1.1), and is often solved with proximal methods.

#### 2.4.4 Spatio-Temporal MRI

Resolving the temporal dynamics of the acquisition during reconstruction yields a wealth of information, such as a time-series images at multiple clinical contrasts and the parametric mapping of the  $T_1$  and  $T_2$  values of the underlying tissue. Additionally, by accounting for the temporal dynamics of the acquisition, the reconstructed images are free from temporal-dynamics related image-blurring [94] as discussed in Section (2.3.3)

However, such applications require three dimensional k-space encoding at each time-point, dramatically increasing the size of the forward model  $A$  while worsening its conditioning, which necessitates better encoding and reconstruction approaches.

This thesis focuses on spatio-temporal MRI applications that leverage partially separable functions as described in [55], which take into account the magnetization temporal dynamics as it varies over the TR.

Let  $m_1, m_2, \dots, m_T$  be the magnetization image as it evolves time, where  $T$  is the number of time points. In the applications considered in this thesis, the imaging process repeats in cycles every TR, resulting in multiple readout lines in a single time point.

If each image is to be resolved using (2.19), the resulting problem would be posed as:

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{bmatrix} = \begin{bmatrix} \mathcal{M}_1 & 0 & \dots & 0 \\ 0 & \mathcal{M}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathcal{M}_T \end{bmatrix} \begin{bmatrix} \mathcal{A}_s & 0 & \dots & 0 \\ 0 & \mathcal{A}_s & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathcal{A}_s \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_T \end{bmatrix} \quad (2.24)$$

where  $\mathcal{A}_s(\cdot) = \mathcal{F}S(\cdot)$

Here,  $i$  denotes the time index, with  $z_i$  representing the multi-channel k-space of image  $m_i$  and  $\mathcal{M}_i$  denoting k-space sampling at that time point. Acquiring enough samples to solve (2.24) is clinically infeasible due to long acquisition times. However, the signal evolutions of realistic  $T_1$  and  $T_2$  values are observed to follow the Bloch equation, and for the tissue parameters within the range of interest, such signal evolution is shown to be well approximated by a low dimensional linear subspace (see Figure (2-3)). An orthonormal basis for this subspace can be incorporated into the forward model.

Let  $\Phi$  be the orthonormal basis of rank  $t_k$  that approximates the signal evolutions. This can then be incorporated into (2.24) as,

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{bmatrix} = \begin{bmatrix} \mathcal{M}_1 & 0 & \dots & 0 \\ 0 & \mathcal{M}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathcal{M}_T \end{bmatrix} \begin{bmatrix} \mathcal{A}_s & 0 & \dots & 0 \\ 0 & \mathcal{A}_s & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathcal{A}_s \end{bmatrix} \Phi \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{t_k} \end{bmatrix} \quad (2.25)$$

$\alpha_i$  are the ‘‘coefficient’’ images that, when passed through  $\Phi$ , will recover the time series  $\{m_i\}$ . In short form, the forward model is described as:

$$z = \mathcal{M}\mathcal{F}S\Phi\alpha \quad (2.26)$$

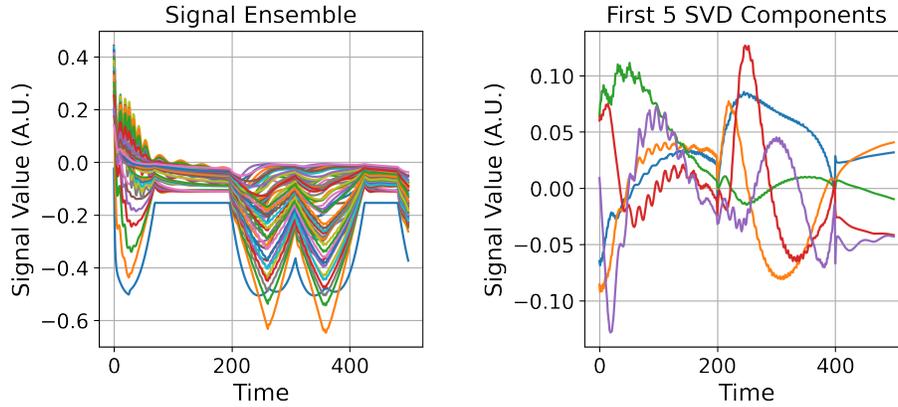


Figure 2-3: Realistic signal evolutions are generated for a particular application (in this case: MRF [61, 15]). The ensemble of signals with different  $T_1$  and  $T_2$  relaxation values is observed to be well approximated by a low-dimensional subspace, which can be calculated using the Singular Value Decomposition (SVD).

Even with the incorporation of  $\Phi$ , applications that leverage (2.26) are often still under-determined, and consequently utilize sampling patterns with spatial and temporal incoherence akin to compressed sensing. A regularized reconstruction (1.1) is then used to recover the underlying coefficient images  $\{\alpha_i\}$ .

$$\operatorname{argmin}_m \frac{1}{2} \|\mathcal{M} \mathcal{F} S \Phi \alpha - y\|_2^2 + \lambda g(x) \quad (2.27)$$

# 3: Polynomial Preconditioning for Regularized Linear Inverse Problems

This chapter has section (2.1) and section (2.2) as “prerequisites”. Section (2.4) is recommended to have better context of the results.

## 3.1: Introduction

In computational Magnetic Resonance Imaging, iterative proximal methods [76] have emerged as the workhorse algorithms to solve linear inverse problems that are posed in the form of (1.1) [59, 9, 60, 58, 34]. This is arguably because:

1. These algorithms are highly generalizable as they do not impose any restrictions on the utilized regularization function ( $g$ ) (as long as the proximal operator of ( $g$ ), defined in the sequel, can be calculated).
2. These algorithms leverage matrix-free implementations of the forward-model  $A$  (i.e., the individual coordinates or entries of  $A$  are not known or stored in memory) for computationally efficient processing of high-dimensional problems.

These reconstructions are typically implemented using the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)[4], which enjoys theoretically optimal convergence. However, the iterative convergence of FISTA and other iterative proximal methods is largely limited by the conditioning of  $A$  as determined by the eigenvalues of  $A^* A$ , as discussed in Lemma 2.2. This limits the integration of MRI applications with an ill-posed  $A$  that leverage (1.1) into clinical practice, as the ill-conditioning results in long reconstruction times.

### 3.1.1 Contributions

This chapter proposes a generalizable polynomial-based preconditioner for faster iterative convergence of regularized linear inverse problems that leverage proximal gradient methods (like FISTA). The evolution of the iterates in proximal gradient descent (PGD) [76] is analyzed and a cost function for polynomial optimization is derived such that the optimized polynomial ( $p$ ), when utilized as a spectral function, directly improves the convergence rate of PGD (and subsequently, FISTA). The proposed preconditioner  $P$  is evaluated as  $P = p(A^* A)$ . Once the polynomial is calculated, it can be applied to any application with forward model  $A$  as long as the maximum eigenvalue of  $A^* A$  can be estimated. (Note that standard FISTA also requires an estimate of the maximum eigenvalue of  $A^* A$ .) The proposed preconditioner does not assume any structure on  $A$ , and can leverage matrix-free implementations of  $A^* A$ . Similarly, the proposed preconditioner does not assume any additional structure on  $g$  other than that its proximal operator, defined in the sequel, can be evaluated. Thus, the proposed preconditioner is highly generalizable and can be applied to various regularized linear inverse problems of interest. In particular, the proposed method retains the “plug-and-play” property of FISTA, and is thus an ideal candidate for integration into reconstruction toolboxes like SigPy [74] and BART [97] whose users typically leverage proximal algorithms in an “out-of-the-box” manner.

### 3.1.2 Related Works

A polynomial-based preconditioner for accelerating the convergence of FISTA, which will be denoted “IFISTA” in this work, was presented in [113]. IFISTA uses a polynomial with binomial coefficients to construct the preconditioner by evaluating the polynomial on  $A^*A$ . In contrast, the polynomial coefficients used in this work are obtained from optimizing a cost function derived from analyzing the error propagation over iterations of PGD which, for a fixed polynomial degree, results in polynomial coefficients that more explicitly target the improved convergence compared to the binomial coefficients used in IFISTA. In fact, the cost function utilized in this work has a natural connection to [46], where a similar cost function was proposed as a means of accelerating the convergence of the Conjugate Gradient. While [46] focuses explicitly on ordinary least squares without regularization and motivates the polynomial design as a means of approximating the inverse of  $A^*A$ , this work arrives at the same cost function for polynomial design, but from the completely different perspective of using PGD to solve ill-posed problems that leverage regularization.

Jacobi-like “left-preconditioning” methods have also been proposed as a means of accelerating convergence, where a matrix  $D$  is designed so that  $A^*DA$  is better conditioned.  $D$  is then incorporated into the DC cost in (1.1) to yield the following [75]:

$$\operatorname{argmin}_x \frac{1}{2} \|D^{1/2}(Ax - b)\|_2^2 + \lambda g(x) \tag{3.1}$$

For non-Cartesian MRI, these methods are typically called “Density Compensation” and have seen wide adoption [79, 84]. Note that by defining  $A_{\text{DCF}}$  and  $b_{\text{DCF}}$  as  $D^{1/2}A$  and  $D^{1/2}b$  respectively, (3.1) is a special case of (1.1). Thus, general methods like FISTA, IFISTA and the proposed preconditioner can synergistically leverage such structure-based left-preconditioning.

In the intersection of computational MRI and regularized linear inverse problems, several preconditioning methods have been proposed that leverage the circulant structure of  $A$ , such as in [51, 86, 105, 66]. In [75], a Frobenius-norm-optimized diagonal preconditioner for the dual variables of the primal-dual hybrid gradient (PDHG) algorithm [19] was presented to improve convergence for non-Cartesian MRI applications. However, by utilizing explicit structure, it is unclear on whether the mentioned methods will generalize well to an arbitrary  $A$ . This increases the barriers to entry of such preconditioning methods, particularly for novel applications. This is particularly true for recent contrast-resolved MRI applications [55, 61, 94, 15, 102, 45, 111, 100, 5, 48, 99, 14] with large measurement sizes and an ill-conditioned forward operator  $A$  such as Echo Planar Time Resolved Imaging [102] and Magnetic Resonance Fingerprinting (MRF) [61, 15, 14]. The high dimensionality and ill-conditioning of  $A$  may also result in computationally intensive procedures to estimate the preconditioner, such as with the Frobenius norm formulation used in [75] and circulant preconditioner design process proposed in [66].

Another advantage of FISTA, IFISTA and the proposed preconditioner is that they can utilize efficient implementations of the normal operator ( $A^*A$ ), which cannot be leveraged by PDHG and [75]. For example, non-Cartesian MRI reconstructions can leverage the Toeplitz structure of the normal operator of the non-uniform Fourier transform to avoid expensive gridding operations [101, 35, 3]. Temporal subspace methods such as  $T_2$ -Shuffling can use the “spatio-temporal” kernel to avoid expanding into the “echo” dimension at each iteration to significantly reduce the number of Fast Fourier Transforms needed [94].

In applications where the regularization ( $g$ ) is cognate to the  $l_1$ -norm to enforce sparsity, algorithms that accelerate Iteratively Re-Weighted Least Squares (IRWLS) [29, 36] have been proposed such as [107] and [20]. However, since these algorithms do not use proximal operators, it limits the different types of regularizations that can be tested without significantly modifying the chosen algorithm. For example, should a user choose to use the algorithm proposed in [20], it is on the user to verify whether their operator  $A$  is “diagonally dominant” and that their chosen regularization ( $g$ ) yields an easy-to-calculate preconditioner (which is re-calculated at every outer iteration).

In the spirit of generalizability and to retain the simplicity and the “plug-an-play” benefits of FISTA, this work compares and contrasts the iterative convergence of the proposed preconditioner against other generalizable methods like FISTA, IFISTA and the Alternating Direction Method of Multipliers (ADMM) [76].

For simplicity, the following analysis focuses on traditional PGD. That being said, the improved con-

ditioning of the method naturally translates into faster convergence when using FISTA as well.

### 3.2: Main Result

Lemma 2.2 motivates the design of a preconditioner that minimizes the magnitude of the eigenvalues of  $I - A^*A$  for faster convergence. The main result of this work is to modify the gradient update steps in Algorithm 2.1 and Algorithm 2.2 using preconditioner  $P = p(A^*A)$ , where  $p$  is designed in such a way that  $p(A^*A)$  increases the contributions of the smaller eigenvalues of  $A^*A$  over iterations for faster convergence.

Let  $p$  have fixed degree  $d$ , where  $d$  is a hyper-parameter that is to be tuned for the application of interest. The coefficients of  $p$  are calculated by optimizing over (3.2), which is motivated in Section (3.3.2).

$$p = \left\{ \underset{q}{\operatorname{argmin}} \int_{z=0}^1 (1 - q(z)z)^2 dz \right. \quad (3.2)$$

The polynomial from (3.2) is then used to derive preconditioner  $p(A^*A)$  that is included in Algorithm 2.1 to arrive at Algorithm 3.1. The derivation is in Section (3.3.1).

#### Algorithm 3.1: Polynomial Preconditioning for PGD

*Inputs:*

- forward model  $A$  →  $y_0 = 0$
- measurements  $b$
- Proximal operator:  $\mathbf{prox}_{\lambda g}$  → Polynomial  $p$  from (3.2)

*Step  $k$ : ( $k \geq 0$ ) Compute*

$$y_{k+1} = \mathbf{prox}_{\lambda g} (y_k - p(A^*A)A^*(Ay_k - b)) \quad (3.3)$$

*Remark.* The iteration variable name has been changed in Algorithm 3.1 (and the following Algorithm 3.2) from  $x_k$  to  $y_k$  to emphasize that the preconditioner is being used.

The preconditioner is integrated into Algorithm 2.2 to get Algorithm 3.2.

#### Algorithm 3.2: Polynomial Preconditioning for FISTA

*Inputs:*

- forward model  $A$  →  $y_{-1} = y_0 = 0$
- measurements  $b$
- Proximal operator:  $\mathbf{prox}_{\lambda g}$  → Polynomial  $p$  from (3.2)

*Step  $k$ : ( $k \geq 0$ ) Compute*

$$y_{k+1} = \mathbf{prox}_{\lambda g} (z_k - p(A^*A)A^*(Az_k - b)) \quad (3.4a)$$

$$\beta_k = k/(k+3) \quad (3.4b)$$

$$z_{k+1} = y_k + \beta_k (y_k - y_{k-1}) \quad (3.4c)$$

### 3.3: Theory

This section will derive and theoretically analyze Algorithm 3.1, which extends to Algorithm 3.2.

#### 3.3.1 Deriving Polynomial Preconditioning for PGD

It helps to motivate Algorithm 3.1 before discussing (3.2). To motivate Algorithm 3.1, first consider the imaging problem posed in noise-free compressed sensing applications [36] with  $b \in \text{range}(A)$ .

$$\begin{aligned} & \underset{x}{\text{argmin}} && g(x) \\ & \text{subject to} && Ax = b \end{aligned} \tag{3.5}$$

In practice, due to measurement errors (such as noise), the following optimization problem is solved instead [36, 59, 9, 58, 34]:

$$\begin{aligned} & \underset{x}{\text{argmin}} && g(x) \\ & \text{subject to} && \|Ax - b\|_2 \leq \epsilon \end{aligned} \tag{3.6}$$

This is often re-cast as the unconstrained problem (1.1) for ease-of-computation. The Karush-Kuhn-Tucker (KKT) conditions can be used to verify that for any  $\epsilon$ , there is an appropriate choice of  $\lambda$  such that the solutions of (3.6) and (1.1) coincide.

Let the singular value decomposition (SVD) of  $A$  in dyadic form be as follows:

$$A(\cdot) = \sum_{i=1}^j \sigma_i \langle \cdot, v_i \rangle u_i \tag{3.7}$$

Here,  $j \leq n, 1 \geq \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_j > 0$  are the singular values of  $A$ ,  $\{u_i\}$  and  $\{v_i\}$  are the left and right singular vectors respectively, and Let  $p(z)$  be a polynomial of degree  $d$  such that  $p(z) > 0$  for  $z \in (0, 1]$ , and let  $P = p(AA^*)$ .

$$P(\cdot) = \sum_{i=1}^j p(\sigma_i^2) \langle \cdot, u_i \rangle u_i \tag{3.8}$$

As an ansatz, let  $P^{\frac{1}{2}}$  be the square root of  $P$ :

$$P^{\frac{1}{2}}(\cdot) = \sum_{i=1}^j [p(\sigma_i^2)]^{\frac{1}{2}} \langle \cdot, u_i \rangle u_i \tag{3.9}$$

#### Lemma 3.1: Injectivity of the Polynomial Preconditioning

The condition  $Ax = b$  in (3.5) is equivalently enforced by the constraint  $P^{\frac{1}{2}}Ax = P^{\frac{1}{2}}b$ .

*Proof.* The condition  $p(z) > 0$  for  $z \in (0, 1]$  implies  $P$  and  $P^{\frac{1}{2}}$  are injective when the domain and co-domain for both operators are restricted to  $\text{range}(A)$ .

Lemma 3.1 motivates the following ‘‘preconditioned’’ formulation that has the same solution as (3.5).

$$\begin{aligned} & \underset{x}{\text{argmin}} && g(x) \\ & \text{subject to} && P^{\frac{1}{2}}Ax = P^{\frac{1}{2}}b \end{aligned} \tag{3.10}$$

To account for model and measurement errors, (3.10) is relaxed to the following constrained formula-

tion:

$$\begin{aligned} & \underset{x}{\operatorname{argmin}} && g(x) \\ & \text{subject to} && \|P^{\frac{1}{2}}(Ax - b)\|_2 \leq \epsilon_p \end{aligned} \quad (3.11)$$

Note that  $\epsilon$  and  $\epsilon_p$  in the constraints of (3.6) and (3.11) respectively are likely to be different as  $P^{\frac{1}{2}}$  is not necessarily unitary.

Similarly to how (3.6) is recast to (1.1), (3.11) is solved by the following for an appropriate choice of  $\lambda_p$ .

$$\underset{x}{\operatorname{argmin}} \quad \frac{1}{2} \|P^{\frac{1}{2}}(Ax - b)\|_2^2 + \lambda_p g(x) \quad (3.12)$$

Defining  $A_p = P^{\frac{1}{2}}A$  and  $b_p = P^{\frac{1}{2}}b$ , the PGD iterations to solve (3.12) and the fixed point condition (Theorem 2.1) of (3.12) is as follows:

$$y_{k+1} = \operatorname{prox}_{\lambda_p g}(y_k - A_p^*(A_p y_k - b_p)) \quad (3.13a)$$

$$y^* = \operatorname{prox}_{\lambda_p g}(y^* - A_p^*(A_p y^* - b_p)) \quad (3.13b)$$

Here,  $y^*$  is the solution to (3.12).

To significantly simplify the iterations of (3.13), the permutability of spectral functions and  $A^*A$ , listed as Lemma 3.2, is leveraged.

#### Lemma 3.2: Permutability of Spectral Functions

$$A^* p(AA^*)A = p(A^*A)A^*A \quad (3.14a)$$

$$A^* p(AA^*) = p(A^*A)A^* \quad (3.14b)$$

*Proof.* Since  $p$  is a spectral function of  $A$ ,  $p(AA^*)$  and  $p(A^*A)$  can be evaluated with respect to the singular value decomposition of  $A$  as defined in (3.7) and (3.8) to verify this result.

Incorporating Lemma 3.2 into (3.13) results in the update step (3.3) of Algorithm 3.1.

$$y_{k+1} = \operatorname{prox}_{\lambda_p g}(y_k - p(A^*A)A^*(Ay_k - b)) \quad (3.15a)$$

$$y^* = \operatorname{prox}_{\lambda_p g}(y^* - p(A^*A)A^*(Ay^* - b)) \quad (3.15b)$$

Extending Lemma 2.2 to (3.15) yields the following.

#### Theorem 3.1: Spectral Interpretation of Polynomial Preconditioning

The error of iterates  $y_k$  in Algorithm 3.1 with respect to  $y^*$  are upper-bounded by the spectrum of  $(I - p(A^*A)A^*A)$ .

*Proof.* Let  $e_k = y_k - y^*$  be the error over iterations. Subtracting the equations of (3.15), taking the  $l_2$ -norm and utilizing Lemma 2.1 yields:

$$\|e_{k+1}\|_2 \leq \|(I - p(A^*A)A^*A)e_k\|_2 \quad (3.16)$$

Similarly to Lemma 2.2, splitting  $e_k$  into  $s_k + t_k$  where  $s_k \in \operatorname{null}(A)$  and  $t_k \in \operatorname{null}(A)^\perp$  results in:

$$\|e_{k+1}\|_2^2 \leq \|(I - p(A^*A)A^*A)t_k\|_2^2 + \|s_k\|_2^2 \quad (3.17)$$

*Remark.* Solving (3.12) with FISTA and using Lemma 3.2 results in Algorithm 3.2.

### 3.3.2 Polynomial Design

Theorem 3.1 motivates finding a polynomial  $p$  such that  $I - p(A^*A)A^*A$  is as close to zero as possible. As the dimensions of  $A$  are typically very large, it is not computationally feasible (in terms of processing time) to perform an eigenvalue decomposition of  $A^*A$  to use as prior information for polynomial design. To avoid this, the coefficients of the polynomial  $p$  is found by optimizing the continuous approximation of the induced norm. This is listed as (3.2). The polynomial that minimizes (3.2) in-turn minimizes the induced  $l_2$ -norm of  $I - p(A^*A)A^*A$  in Theorem 3.1. In other words, the component  $t_k$  in  $v_i$  (as in (3.17) and (3.7)) is upper-bounded by  $|1 - p(\sigma_i^2)\sigma_i^2|$ , which according to (3.2), is minimized to be close to zero. Note that the larger degree  $d$  of polynomial  $p$ , the better  $(1 - p(x)x)^2$  (equivalently  $|1 - p(x)x|$ ), can approximate the zero function.

Priors on the spectrum can be easily incorporated into (3.2) as follows:

$$p = \left\{ \operatorname{argmin}_q \int_{z=0}^1 w(z)(1 - q(z)z)^2 dz \right. \quad (3.18)$$

Here,  $w$  can weight the cost to prioritize certain components of the spectrum.

At first glance, minimizing the following objective instead of (3.2) is preferable as it directly translates into minimizing the appropriate induced norm of Theorem 3.1:

$$p = \left\{ \operatorname{argmin}_q \max_{z \in (0,1)} |1 - q(z)z| \right. \quad (3.19)$$

It is well known that Chebyshev polynomials of the first kind can be used to derive the optimal polynomials such that the maximum absolute value of that polynomial over a specified interval is minimized. However, defining  $r(z) = 1 - q(z)z$ , and using Chebyshev polynomials to determine  $r$  yields a polynomial with  $r(z) = 1$  for multiple values of  $z \in [0, 1]$  due to the constraint  $r(0) = 1$ , which implies the components of  $t_k$  in (3.17) corresponding to eigenvalues  $\sigma^2$  such that  $r(\sigma^2) = 1$  will not decrease. However, if the minimum non-zero eigenvalue  $\mu = \sigma_j^2$  of  $A^*A$  is known a-priori, the following polynomial minimizes (3.19) over the interval  $[\mu, 1]$  (See [47, 98, 91, 46]):

$$r(z) = \frac{T_{d+1}\left(\frac{1+\mu-2z}{1-\mu}\right)}{T_{d+1}\left(\frac{1+\mu}{1-\mu}\right)} \quad (3.20a)$$

$$p(z) = \frac{1 - r(z)}{z} \quad (3.20b)$$

Here,  $T_{d+1}$  is the Chebyshev polynomial of the first kind of degree  $d + 1$ . In practice, (3.2) is preferred as it is often computationally expensive to estimate  $\mu$  unless  $A^*A$  happens to be injective, in which case  $\mu$  can be estimated by performing power-iteration on  $I - A^*A$ .

Rather interestingly, while (3.2) and (3.19) were motivated by studying the evolution of iterates in PGD when solving regularized linear inverse problems in the form of (1.1), the exact formulation for polynomial optimization was studied in [46] as a means of accelerating the convergence of Conjugate Gradient for ordinary least squares optimization, where (3.2) and (3.19) were optimized to construct an incomplete inverse of  $(A^*A)$  to use a preconditioner.

Theorem 3.2 verifies that the positivity assumption in Lemma 3.1 is satisfied by optimized results of (3.2) and (3.19).

### Theorem 3.2: Polynomial Positivity

The polynomial  $p$  derived from optimized (3.2) and (3.19) satisfies the constraint  $p(z) > 0$  for  $z \in (0, 1]$ .

*Proof.* This is a consequence of Theorem 4 of [46].

### 3.3.3 Error Bound

With the simplifying assumptions that  $A^*A$  is injective and  $\lambda_p = \lambda$ , Theorem 3.3 bounds the difference between the solutions of (1.1) and (3.12). Let  $\|\cdot\|_{2 \rightarrow 2}$  denote the induced  $l_2$ -norm.

#### Theorem 3.3: Bounding Error when using Injective Left-Preconditioner

Assume  $A^*A$  is injective, and  $\lambda = \lambda_p$  in (1.1) and (3.12) respectively. Let  $e_k, \gamma$  and  $\delta$  be defined as:

$$e_k = y_k - x^* \quad (3.21a)$$

$$\gamma = \|I - p(A^*A)A^*A\|_{2 \rightarrow 2} \quad (3.21b)$$

$$\delta = \|(I - p(A^*A))A^*(Ax^* - b)\|_2 \quad (3.21c)$$

Here,  $x^*$  is the optimal solution to (1.1),  $y_k$  are the iterates of Algorithm 3.1 that leverages polynomial  $p$ . Then, the limit of the difference,  $e_\infty$ , satisfies the following:

$$\|e_\infty\|_2 \leq \frac{\delta}{1 - \gamma} \quad (3.22)$$

*Proof.* Subtracting (3.3) from (2.6), taking the  $l_2$ -norm of the difference and utilizing Lemma 2.1 with the triangle inequality yields:

$$\|e_{k+1}\|_2 \leq \|(I - PA^*A)e_k\|_2 + \|(I - P)A^*(Ax^* - b)\|_2 \quad (3.23)$$

This reduces to:

$$\|e_{k+1}\|_2 \leq \gamma \|e_k\|_2 + \delta \quad (3.24)$$

As  $A^*A$  is injective, it follows that, after optimizing for  $P$  via (3.2) or (3.19),  $\gamma < 1$ . This in-turn implies (3.24) converges and the limit  $e_\infty$  satisfies (3.22).

For most regularized linear inverse problems,  $\|Ax^* - b\|$  is small, which in turn implies the error between (1.1) and (3.12) is also small as:

$$\delta \leq \|I - p(A^*A)\|_{2 \rightarrow 2} \|A^*(Ax^* - b)\|_2 \quad (3.25)$$

### 3.3.4 Implementation Details and Complexity Analysis

Prior works [113, 46] that used polynomial preconditioning utilized the matrix entries of  $A$  to explicitly pre-calculate  $p(A^*A)$  to save computation time. Since this work aims to leverage matrix-free implementations of  $A$ , this is no longer possible. Rather, utilizing a polynomial preconditioner of degree  $d$  involves  $(d + 1)$   $A^*A$  evaluations per iteration, thereby increasing the per-iteration cost compared to PGD and FISTA. However, the main benefit of the preconditioner is that, for the same number of  $A^*A$  evaluations, the components of the iterates corresponding to the smaller eigenvalues of  $A^*A$  are more explicitly targeted. To make sure this point is well reflected in the sequel, the experiments in Section (3.4) reports convergence as a function of the total number of  $A^*A$  evaluations (and the number of iterations, which

reflects the number of proximal operator evaluations) in addition to reporting the iterative convergence as a function of wall time to demonstrate real world performance.

To give an example, Figure (3-1) depicts the iteration progress assuming a degree 3 polynomial optimized using (3.2) after 4  $A^*A$  evaluations for Algorithm 2.1 and Algorithm 3.1, which shows that the polynomial preconditioner aids the convergence of components corresponding to the lower eigenvalues of  $A^*A$  at the cost of slightly slower convergence of the higher eigenvalues. In the sequel, the degree of the polynomial is a hyper-parameter that is tuned for the application of interest.

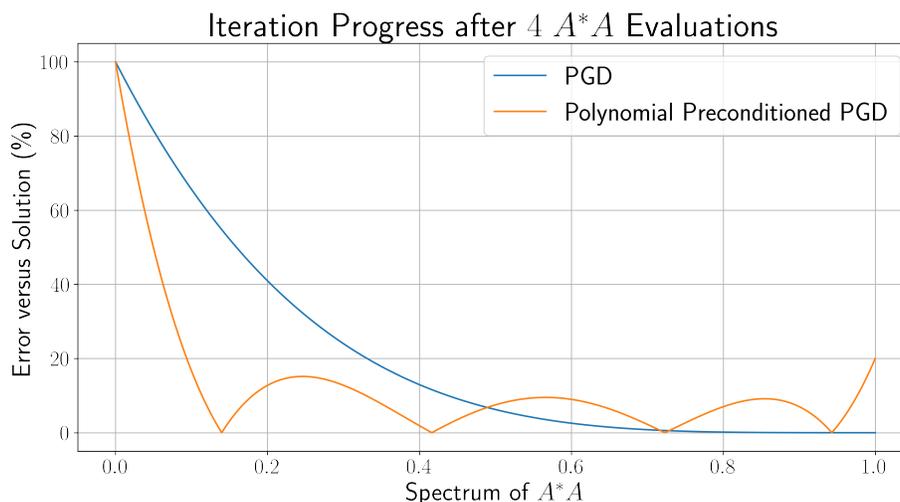


Figure 3-1: Iteration progress of Algorithm 2.1 and Algorithm 3.1 assuming a degree 3 polynomial optimized using (3.2) after 4  $A^*A$  evaluations. The x-axis denotes the spectrum of  $A^*A$ . The y-axis denotes the percentage of decrease of error with respect to the respective final iterates. This figure shows that the polynomial preconditioner aids the convergence of components corresponding to the lower eigenvalues of  $A^*A$  at the cost of slower convergence of the higher eigenvalues.

Note that for both, numerical stability and achieving  $(d + 1)$  evaluations of  $A^*A$  per iteration, it is important to leverage the polynomial structure to reduce the number of  $A^*A$  calls when evaluating the preconditioner. For example, consider pseudo-code implementation in Listing 3.1 that utilizes recursion. It is assumed that “ListOfPolynomialCoefficients” returns the coefficients of  $p$  in the order:

$$p(x) = \sum_{i=0}^d c[i]x^i \quad (3.26)$$

### 3.3.5 Interpretation and Noise Coloring

At first glance, the proposed preconditioner is similar to the left-preconditioning methods such as (3.1). However, instead of using a diagonal matrix to weight to the entries (or coordinates) of the measurement  $b$ ,  $P^{1/2}$  “spectrally” weights  $b$ . In other words,  $P^{1/2}$  is a diagonal matrix with respect to the left singular vectors of  $A$  (or  $\{u_i\}$ ). A unique consequence of this property is that, in the absence of regularization, polynomial preconditioning result in the same solution as (1.1) and thus introduces no noise coloring. This is not necessarily true for general left-preconditioning methods.

#### Theorem 3.4: Identical $l_2$ Solutions Without Regularization

When  $\lambda = \lambda_p = 0$  in (1.1) and (3.12) respectively, the solutions to (1.1) and (3.12) are identical as long as both optimizations are equivalently initialized.

Listing 3.1: Pseudo-code for Polynomial Evaluation

```

I = IdentityOperator ()
N = ForwardModelNormalOperator ()

def CreatePolynomialPreconditioner (c):
    if len(c) == 1:
        return c[0] * I
    else:
        return c[0] * I + N * CreatePolynomialPreconditioner (c[1:])

coeffs = ListOfPolynomialCoefficients ()
Preconditioner = CreatePolynomialPreconditioner (coeffs)

```

*Proof.* Using (3.7) and (3.8),

$$\begin{aligned}
 x \text{ solves (1.1)} &\iff \sigma_i \langle x, v_i \rangle = \langle b, u_i \rangle \\
 &\iff p(\sigma_i^2)^{\frac{1}{2}} \sigma_i \langle x, v_i \rangle = p(\sigma_i^2)^{\frac{1}{2}} \langle b, u_i \rangle \\
 &\iff x \text{ solves (3.12)}
 \end{aligned} \tag{3.27}$$

### 3.4: MRI Experiments

In the spirit of reproducible research, the data and code used to perform the following experiments can be found at:

<https://github.com/sidward/ppcs><sup>1</sup>

All reconstructions were implemented in the Python programming language using SigPy<sup>2</sup>[74]. The polynomial optimizations (3.2) were performed using SymPy<sup>3</sup> [64], with the latter leveraging an excellent Chebyshev polynomial package available at <https://github.com/mlazaric/Chebyshev><sup>4</sup>.

To verify the efficacy of the preconditioner, three varied MRI reconstructions were studied using the unconstrained formulations (1.1) and (3.12). The Nesterov accelerated variants, Algorithm 2.2 and Algorithm 3.2, were used over their respective non-accelerated counterparts. For all cases, the corresponding measurement matrix  $A$  was normalized to have a unitary induced  $l_2$ -norm, and the measurement vector  $b$  was normalized to have unitary  $l_2$ -norm. All experiments were performed on an Intel (R) Xeon Gold 5320 CPU and an NVIDIA(R) RTX A6000 GPU.

#### 3.4.1 Parameter Selection

Before discussing the specific experiments, this section will describe the efforts undertaken to ensure that each algorithm is portrayed in the best light.

Solving linear inverse problems posed as (1.1) and (3.12) using FISTA, IFISTA, ADMM and the proposed method requires the user to specify the regularization value (for each algorithm) and the desired number of  $A^*A$  evaluations per iteration (for each algorithm modulo FISTA). FISTA utilizes one  $A^*A$  evaluation per iteration. IFISTA and the proposed polynomial preconditioner, when using a polynomial  $d$ , utilizes  $(d+1)$   $A^*A$  evaluations per iteration. ADMM utilizing  $n$   $A^*A$  evaluations per iteration implies  $(n-1)$  conjugate gradient (CG) iterations during the inner loop of ADMM (as one  $A^*A$  evaluation is needed to

<sup>1</sup><https://doi.org/10.5281/zenodo.6475880>

<sup>2</sup><https://doi.org/10.5281/zenodo.5893788>

<sup>3</sup><https://doi.org/10.7717/peerj-cs.103>

<sup>4</sup><https://doi.org/10.5281/zenodo.5831845>

calculate the initial residual term in the CG algorithm). IFISTA and the proposed method require additional tuning of the regularization value as argued by the introduction of  $\lambda_p$  in (3.12).

The convergence of the iterates of the respective algorithms and the quality of the converged result are dependent on the regularization value chosen and the number of  $A^*A$  evaluations per iteration. A fair comparison between the algorithms necessitates a search through the large parameter space of regularization values and the number of normal evaluations. It is difficult to manually choose parameters based on the observed convergence and reconstruction quality. To overcome this, the following programmatic procedure is used to select the parameters.

1. For all algorithms, the maximum number of total  $A^*A$  evaluations is fixed.
2. The minimum normalized root mean squared error (NRMSE) achieved by FISTA (with respect to the reference image) among all tested regularization values ( $\lambda$ ) is noted as  $\epsilon_f\%$ .
3. For each method, if a parameter set yields a reconstruction with NRMSE (with respect to the reference image) that is greater than  $(\epsilon_f + 2)\%$ , then that parameter set is discarded. The “+2” term was set to account for the finite resolution of the grid search of regularization values ( $\lambda$ ) and potential numerical errors. This “+2” term was observed to yield qualitatively similar reconstruction to the minimum NRMSE FISTA result for each of the following experiments. Note that the reconstruction error achieved might be lower than  $\epsilon_f\%$ .
4. For each method, the convergence curves (with respect to the last iterate) as a function of the observed wall-time is calculated for the remaining parameter sets that satisfy the  $(\epsilon_f + 2)\%$  constraint.
5. A log-linear fitting of the convergence curves is performed, and the parameters with the most negative slope (after the log-linear fitting) is chosen to represent the method.

By selecting parameters that yield reconstructions that fall within  $(\epsilon_f + 2)\%$  NRMSE of the reference, the reconstruction quality is assured. Given the subset of parameters, the convergence is calculated with respect to the last iterate of the respective methods instead of the reference image to directly show-case iterative convergence, which allows the resulting plots in the following experiments to be directly interpreted in the context of equations (2.9) and (3.17).

Please see Section (3.4.2) for a visual example of the parameter selection process proposed in this section.

### 3.4.2 Cartesian MRI

The first experiment is a 2D-Cartesian Compressed Sensing Knee application using publicly available data [73, 89]. The reference data was acquired using a 3D-FSE CUBE acquisition with proton density weighting that included fat saturation [89] on a 3T whole-body scanner (GE Healthcare, Waukesha, WI) using an 8-channel HD knee coil (GE Healthcare, Milwaukee, WI, USA) with an echo time (TE) of 25ms and repetition time (TR) of 1550ms. The field-of-view was 160mm, the matrix size was  $320 \times 320$ , slice thickness was 0.6mm and 256 slices were acquired. The reference data was fully-sampled and satisfied the Nyquist criterion. This reference data was retrospectively under-sampled by approximately  $R = 7.21$  times using a variable density Poisson disc sampling mask generated by BART [97].

The unconstrained reconstruction formulation (1.1) for this experiment is as follows:

$$x^* = \left\{ \underset{x}{\operatorname{argmin}} \frac{1}{2} \|MFSx - b\|_2^2 + \lambda \|Wx\|_1 \right. \quad (3.28)$$

Here,  $W$  is the forward Daubechies-4 Wavelet transform,  $S$  is the SENSE model of the parallel-imaging acquisition [85] estimated using [96],  $F$  is the 2D-Fourier transform and  $M$  is the Poisson disc sampling mask.

This application was solved with FISTA, IFISTA, ADMM and FISTA with the polynomial preconditioner. Each algorithm was allowed to run for a maximum of 60  $A^*A$  evaluations. For FISTA, IFISTA and

polynomial preconditioned FISTA, regularization values  $\lambda$  were varied as:

$$\lambda \in \left\{ \frac{10^{-2}}{1.5^k} : k = 0, 1, \dots, 14 \right\} \quad (3.29)$$

For IFISTA, ADMM and the polynomial preconditioned FISTA, the number of  $A^*A$  evaluations were varied as:

$$A^*A \text{ evaluations per iteration} \in \{2, 3, 4, 5, 6\} \quad (3.30)$$

The number of  $A^*A$  evaluations per iteration were chosen to cleanly divide the maximum number of  $A^*A$  evaluations for fair comparison between methods. Lastly, the regularization value from the FISTA experiments with the least NRMSE against the reference image was used for ADMM, with the step size ( $\rho$ ) varied as:

$$\rho \in \left\{ 3^k : k = -7, -6, \dots, -1, 0, 1, \dots, 6, 7 \right\} \quad (3.31)$$

To give a visual example of how the programmatic selection works (Section (3.4.1)), among the  $\lambda$  values tested in (3.29), the minimum NRMSE achieved by FISTA was  $\epsilon_f = 13.61\%$ . It is observed that for FISTA, four different regularization values  $\lambda$  achieved ( $\epsilon_f + 2$ )% or less NRMSE with respect to the reference image. Qualitatively, the  $\lambda = 1.73 \times 10^{-4}$  curve achieved the fastest iterative convergence as shown in Figure (3-2)(A). This is seen to translate to the most negative slope for the corresponding log-linear fit as shown in Figure (3-2)(B) and is thus the parametric value chosen to represent FISTA.

For IFISTA, ADMM and the proposed method, the identical process was used (with the same ( $\epsilon_f + 2$ )% NRMSE condition) to choose representative regularization value and number of  $A^*A$  evaluations per iteration.

The convergence curves of the respective methods after the programmatically chosen parameters with respect to both, the number of  $A^*A$  evaluations and observed wall times, are depicted in Figure (3-3), and the final iterates of the corresponding methods are depicted in Figure (3-4)

### 3.4.3 Non-Cartesian MRI

The second experiment is a 2D non-Cartesian variable-density spiral brain application using data that is publicly available at the code repository<sup>5</sup>. Reference data was acquired on a 3T scanner (GE Healthcare, Waukesha, WI) with IRB approval and informed consent obtained using 32 coils with a variable density spiral trajectory at  $1\text{mm} \times 1\text{mm}$  resolution and a field-of-view of  $220\text{mm} \times 220\text{mm}$ . The trajectory was designed with 16-fold in-plane under-sampling at the center of k-space and a linearly increasing under-sampling rate up to 32-fold at the edge of k-space. The readout duration was 6.7ms. This reference data was SVD coil-compressed to 12 coils, and then retrospectively under-sampled by discarding every other interleaves out of 32 interleaves.

The unconstrained reconstruction formulation (1.1) for this experiment is as follows:

$$x^* = \left\{ \underset{x}{\operatorname{argmin}} \frac{1}{2} \|\mathcal{F}Sx - b\|_2^2 + \lambda \|Wx\|_1 \right. \quad (3.32)$$

Here,  $W$  is the forward Daubechies-4 Wavelet transform,  $S$  is the SENSE model of the parallel-imaging acquisition [85] estimated using [96], and  $\mathcal{F}$  is the non-uniform Fourier transform. This experiment utilized the Toeplitz structure of  $\mathcal{F}^* \mathcal{F}$  for faster evaluation for both (1.1) and (3.12)[101, 35, 3].

This application was solved with FISTA, IFISTA, ADMM and FISTA with the polynomial preconditioner. The regularization values ( $\lambda, \rho$ ) tested were the same as in Section (3.4.2). The maximum number of  $A^*A$  set to 80. For IFISTA, ADMM and the polynomial preconditioned FISTA, the number of  $A^*A$  evaluations were varied as:

$$A^*A \text{ evaluations per iteration} \in \{2, 4, 5, 8, 10\} \quad (3.33)$$

The number of  $A^*A$  evaluations per iteration were chosen to cleanly divide the maximum number of  $A^*A$  evaluations for fair comparison between methods.

<sup>5</sup><https://github.com/sidward/ppcs>

The same programmatic procedure as in Section (3.4.1) was used to select representative parameters for each method with the best FISTA reconstruction achieving an NRMSE of  $\epsilon_f = 16.03\%$ . The convergence curves of the respective methods after the programmatically chosen parameters with respect to both, the number of  $A^*A$  evaluations and observed wall times, are depicted in Figure (3-5), and the final iterates of the corresponding methods are depicted in Figure (3-6). Figure (3-7) shows the utilized trajectory in (3.32).

### 3.4.4 Spatio-Temporal MRI

The third experiment is a 3D non-Cartesian spatio-temporal brain application. The data were obtained on a 3T scanner (GE Healthcare, Waukesha, WI) with IRB approval and informed consent obtained. The reference data were acquired with 48 channel coils using a 3D MRF[61] acquisition with the TGAS-SPI trajectory with a total of 48 groups acquired to achieve adequate 3D k-space encoding at each temporal data point [15].

Each acquisition group contains an adiabatic inversion preparation with TI of 15ms and 500 variable flip-angle acquisitions with TR of 12.5 ms and TE of 1.75ms, with a 1.2s wait time for signal recovery to improve the signal-to-noise ratio, resulting in a net acquisition per group of 7.45s and total acquisition time of approximately 6 minutes. Additionally, a water-exciting rectangular pulse with duration of 2.38ms was used to depress the fat signal [72]. A variable density spiral trajectory with a 16-fold in-plane under-sampling rate at the center of k-space and a 32-fold under-sampling rate at the edge of k-space was used to achieve an encoding at 1-mm isotropic resolution with a field-of-view of  $220 \times 220 \times 220 \text{ mm}^3$  with a readout duration of 6.7 ms. This reference data was coil-compressed to 8 coils (from 48) using a combination of [49] and SVD coil compression. Retrospective under-sampling was performed to simulate a 1 minute acquisition. The data used is publicly available at the code repository<sup>6</sup>.

The unconstrained reconstruction formulation (1.1) for this experiment is as follows [55, 78, 62, 112, 94, 15]:

$$x^* = \left\{ \underset{x}{\operatorname{argmin}} \frac{1}{2} \|D^{1/2} (\mathcal{F}S\Phi x - b)\|_2^2 + \lambda \text{LLR}(x) \right. \quad (3.34)$$

Here, LLR is the locally-low rank constraint [94] with a block-size of  $8 \times 8 \times 8$ ,  $S$  is the SENSE model of the parallel-imaging acquisition [85] estimated using [109] and  $\mathcal{F}$  is the 3D non-uniform Fourier transform. Note that in this case, the reconstructed  $x$  consists of multiple ‘‘coefficient’’ images such that  $\Phi x$  recovers the temporal evolution of the underlying signal. Please see [55, 61, 94, 15, 102, 45, 111, 100, 5, 99] for more information. (3.34) leverages structural left-preconditioning  $D$  for faster convergence. The particular method used to derive  $D$  was [79]. Since the structural left-preconditioner ( $D$ ) derived from [79] differs between the 6-minute acquisition and 1-minute acquisition, the 1-minute FISTA reconstruction of the data is used as a reference for fair comparison.

The inclusion of  $D$  reduces the theoretical efficacy of the polynomial preconditioner as the resulting normal operator when solving (3.34) (i.e.  $A^*DA$ ) has a much narrower eigenvalue spectrum compared to  $A^*A$ . That being said, utilizing the polynomial preconditioner (which can be directly applied to (3.34) thanks to its generalizability) can still be beneficial to reduce real world reconstruction times. In particular, given the same number of  $A^*A$  evaluations, the polynomial preconditioner achieves qualitatively similar reconstruction to the non polynomial preconditioned result while utilizing fewer proximal calls. For applications where evaluating  $A^*A$  is much faster than a proximal call (such as in the sequel), the fewer proximal evaluations yields significant reductions in processing times.

For this TGAS-SPI-MRF application, (3.34) is too big to solve directly on even high-end GPUs without utilizing some form of ‘‘batching’’ that moves data between CPU memory and GPU memory when evaluating matrix-vector products of  $A$  and  $A^*$ . This increases the per-iteration costs. Instead, an ADMM formulation is leveraged to split (3.34) into two smaller sub problems, where each sub problem completely fits (that is to say, there is no need to implement any kind of batching) in approximately 20GB of GPU memory. Each sub problem can then be solved in parallel on separate GPU devices with significantly fewer data transfers between CPU and GPU devices.

Let  $(A_1, b_1)$  denote the forward model and acquired data of the first four (of eight) receiver channels,

<sup>6</sup><https://github.com/sidward/ppcs>



variables or an application-specific preconditioner array, resulting in comparable computational memory requirements to FISTA. Additionally, it enables the use of the faster  $A^*A$  evaluations if applicable, such as in Section (3.4.3), which is not possible with [75].

While the proposed preconditioner does increase the per-iteration computational cost compared to FISTA, the proposed method enables faster computation compared to FISTA even when requiring that both methods to utilize an equal number of  $A^*A$  evaluations. This is because, on-top of the theoretically faster convergence offered by the polynomial preconditioner (as verified by the convergence curve as a function of the number of  $A^*A$  evaluations), utilizing the preconditioner enables fewer proximal operator evaluations while still achieving comparable reconstruction. In particular, for the experiments showcased in this manuscript, the proximal operator is more computationally expensive compared to evaluating  $A^*A$ , translating to significantly faster real-world performance as demonstrated by the convergence curve versus wall-time, while achieving similar reconstruction performance. The reduced number of proximal evaluations of the proposed method will be of even more benefit for highly intensive proximal operators, such as the structured matrix completion approaches akin to the soft-thresholding versions of [92, 40].

The generalizability of the polynomial preconditioner allows it to be used directly in a subspace reconstruction [55, 94, 15, 78, 62, 112] without needing to explicitly account for  $\Phi$  when designing the preconditioner while also easily integrating the structure-specific left-preconditioning ( $D$  in (3.34)). By utilizing  $A^*A$  to construct the preconditioner, the method inherently takes into account information from  $\Phi, S, \mathcal{F}$  and  $\mathcal{D}$  without user modification. That being said, the amount of the theoretical convergence benefits of the proposed preconditioner is dependent on the spectrum of  $A^*A$ .

Looking at the convergence of iterates as a function of the number of  $A^*A$  evaluations, IFISTA performs slightly worse than standard FISTA in Section (3.4.2) and Section (3.4.3). This is arguably because, as discussed in Section (3.3.4), the IFISTA preconditioner is instead evaluated in a matrix-free manner. In contrast, by explicitly optimizing for the faster convergence of iterates for a given degree via (3.2), the proposed polynomial preconditioner enables faster convergence.

This work does not explore incorporating a weighting-prior into the spectral cost ( $w$  in (3.18)). It is expected that a reasonable prior estimate of the spectrum of an operator  $A$  derived for a specific application will significantly improve the rate of convergence for that application. However, given the large dimensionality of  $A$  and the consequent difficulty in approximating the spectrum of  $A^*A$ , estimating a reasonable prior  $w$  will involve trial-and-error, and is thus left to future work.

A limitation of the polynomial preconditioner is with respect to numerical stability. In principle, it is possible to utilize a polynomial  $p$  of a high degree  $d$ . However, in practice, evaluating powers of  $A^*A$  can accumulate numerical errors. Therefore, tuning the degree  $d$  for the application of interest is required.

The error bound presented in Theorem 3.3 is applicable to any linear operator  $P$  such that  $P$  is injective on  $\text{null}(A)^\perp$ . Thus, deriving a  $P$  to minimize the error bound in (3.22) is a promising avenue for application-specific preconditioner design. For example, (3.22) can be used to upper-bound the error of the circulant preconditioner in [66].

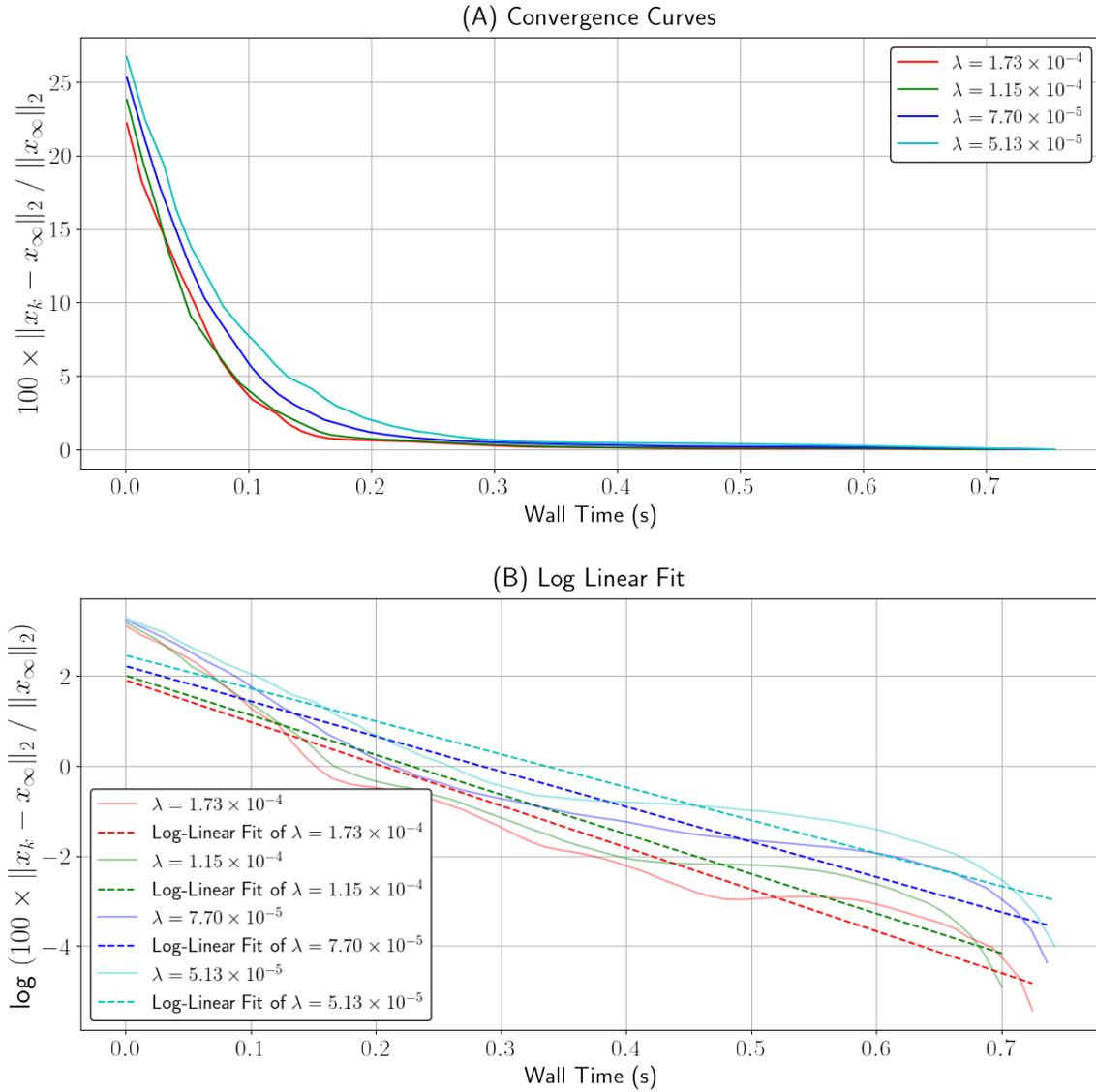
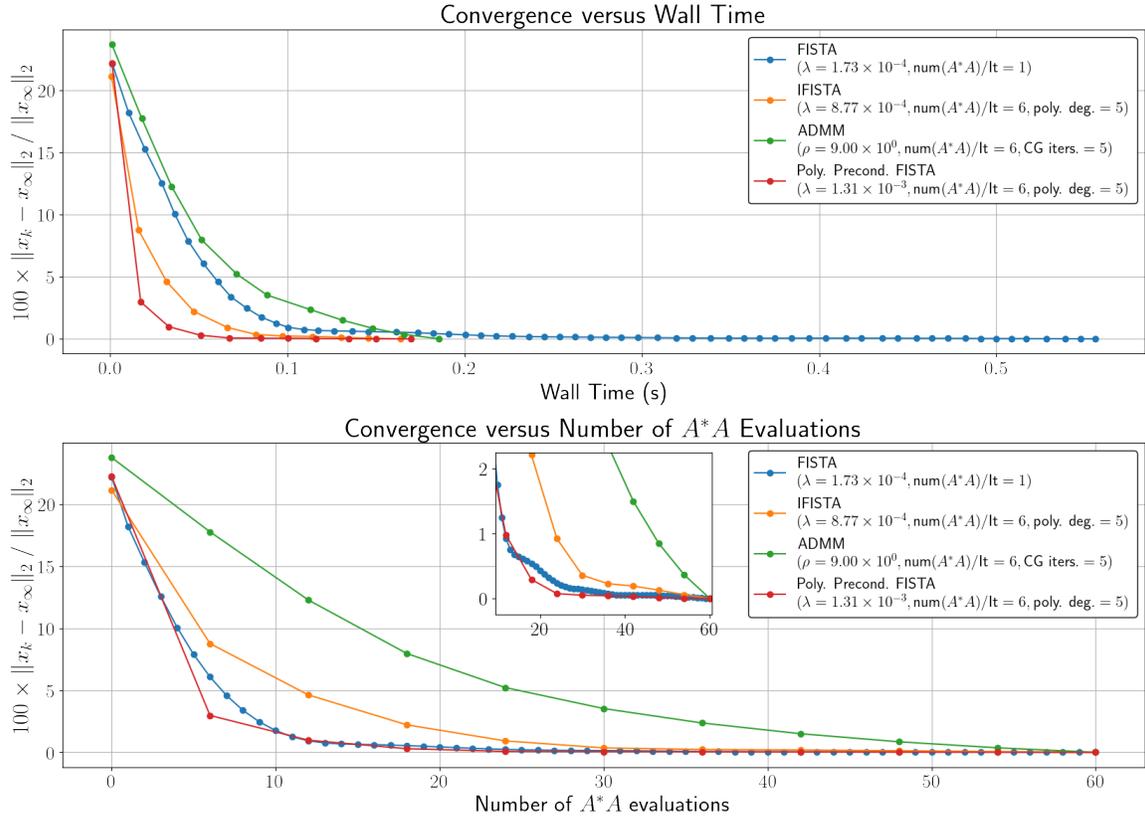


Figure 3-2: This figure demonstrates how the hyper-parameters were chosen to represent FISTA. Among the hyper-parameters tested in (3.29), the minimum NRMSE against the reference image achieved by FISTA was  $\epsilon_f = 13.61\%$ . The four  $\lambda$  values shown in (A) and (B) achieved  $(\epsilon_f + 2)\%$  NRMSE. In order to programmatically pick the fastest converging value between the listed four values, a log-linear fitting of the convergence curves is performed, as shown in (B). The parameter with the most negative log-linear slope is chosen to represent the algorithm, which in this case is the  $\lambda = 1.73 \times 10^{-4}$  curve.



**Figure 3-3: Cartesian MRI Convergence Results.** This figure depicts the convergence results of the respective methods described in Section (3.4.2) given the hyper-parameters selected as in Section (3.4.1) with  $\epsilon_f = 13.61\%$ . Given the chosen hyper-parameters, the error over iterations with respect to the last iteration of each respective method is plotted. The  $k^{\text{th}}$  iteration and last iteration are labelled as  $x_k$  and  $x_\infty$  respectively. The legend on the top right depicts the algorithm and the chosen hyper-parameters. The x-axis of the top and bottom subplots denotes the total number of  $A^*A$  evaluations and measured wall-times respectively. The circular-markings on each line denote the respective iteration points. The number of proximal operators evaluated by a point on the x-axis is equal to the number of iterations by that point.

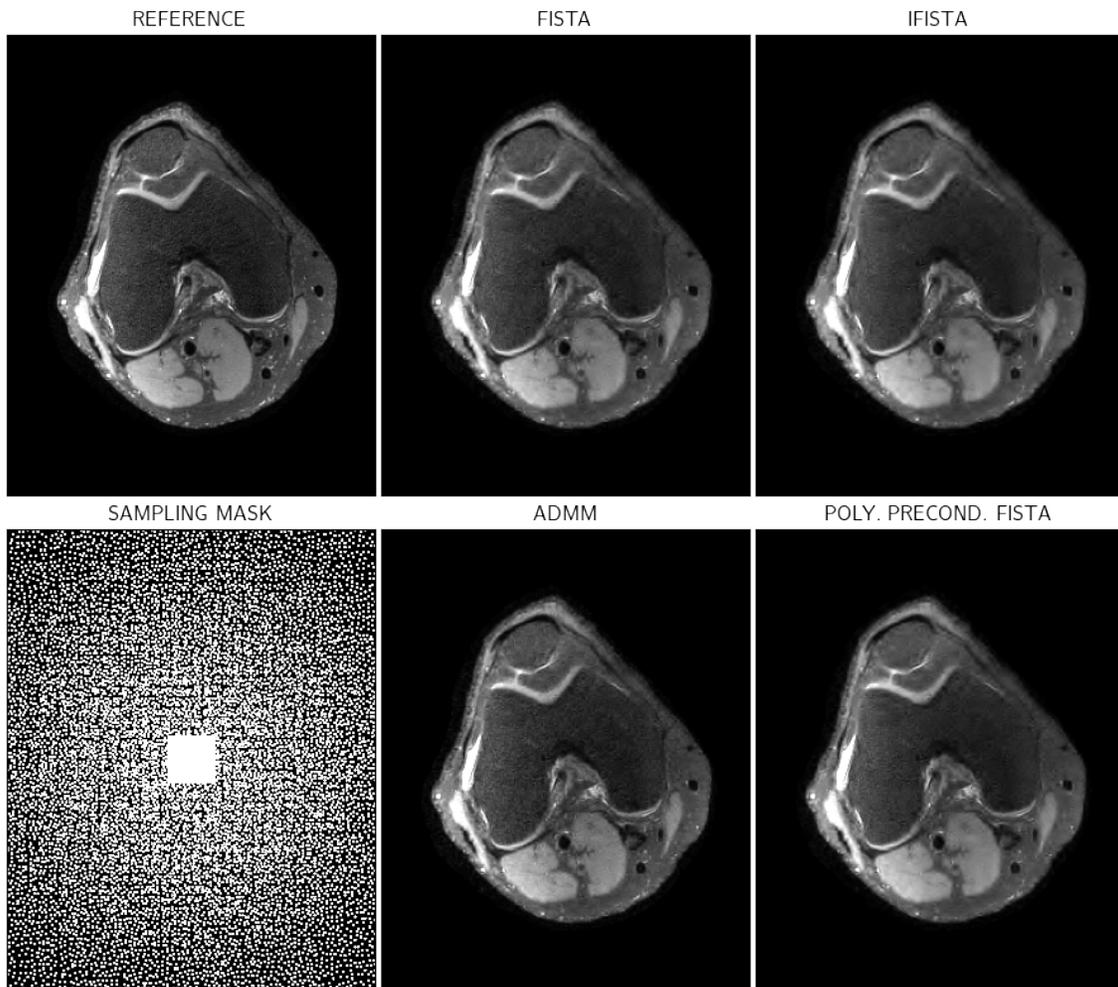
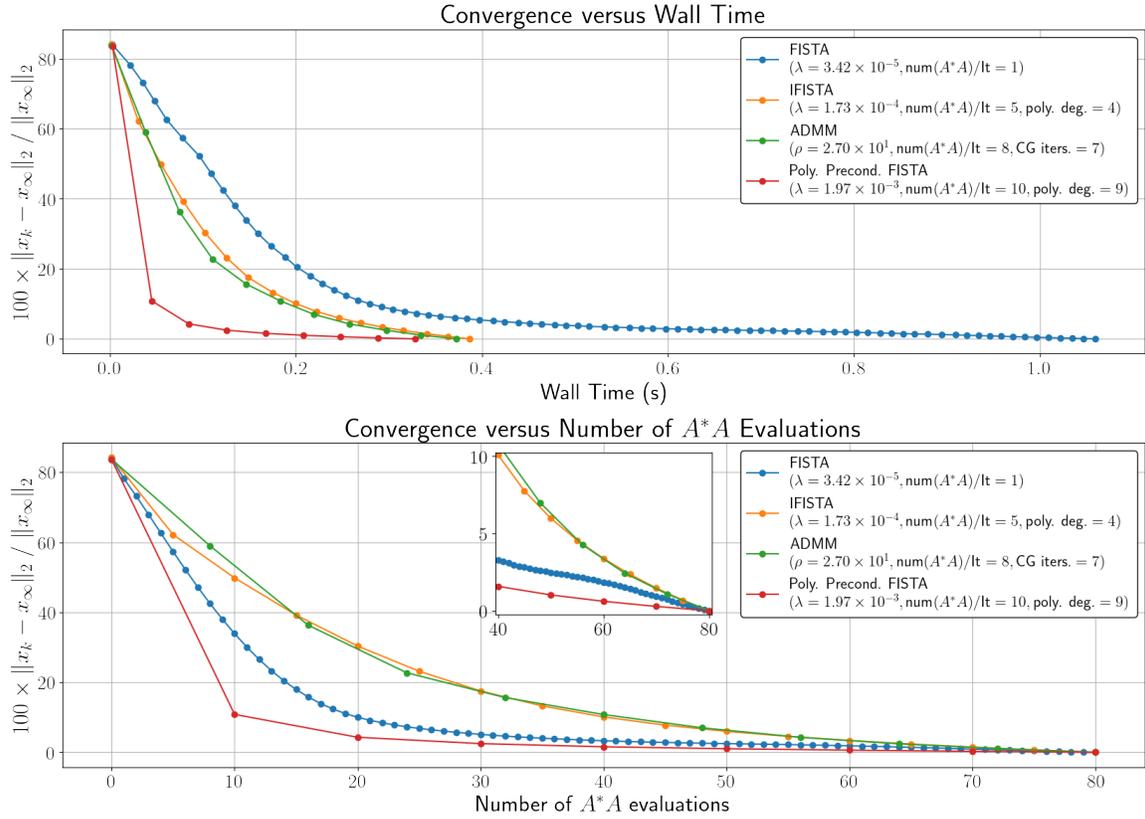


Figure 3-4: **Cartesian MRI Reconstruction Results.** This figure depicts the final iterations of the respective methods in figure (3-3) The bottom left figure, labelled “SAMPLING MASK”, denotes the under-sampling mask used in (3.28). The hyper-parameters for these results are depicted in figure (3-3).



**Figure 3-5: Non-Cartesian MRI Convergence Results.** This figure depicts the convergence results of the respective methods described in Section (3.4.3) given the hyper-parameters selected as in Section (3.4.1) with  $\epsilon_f = 16.03\%$ . Given the chosen hyper-parameters, the error over iterations with respect to the last iteration of each respective method is plotted. The  $k^{\text{th}}$  iteration and last iteration are labelled as  $x_k$  and  $x_\infty$  respectively. The legend on the top right depicts the algorithm and the chosen hyper-parameters. The x-axis of the top and bottom subplots denotes the total number of  $A^*A$  evaluations and measured wall-times respectively. The circular-markings on each line denote the respective iteration points. The number of proximal operators evaluated by a point on the x-axis is equal to the number of iterations by that point.

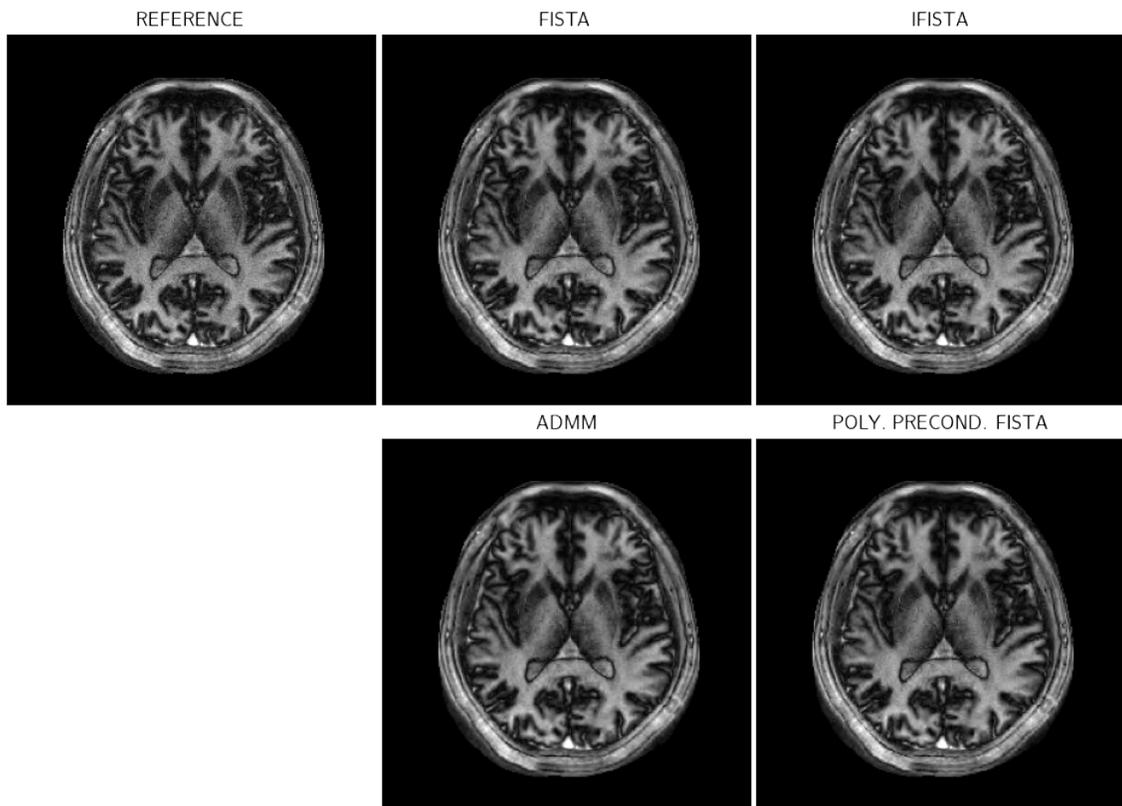


Figure 3-6: **Non-Cartesian MRI Reconstruction Results.** This figure depicts the final iterations of the respective methods in figure (3-5). The hyper-parameters for these results are depicted in figure (3-5).

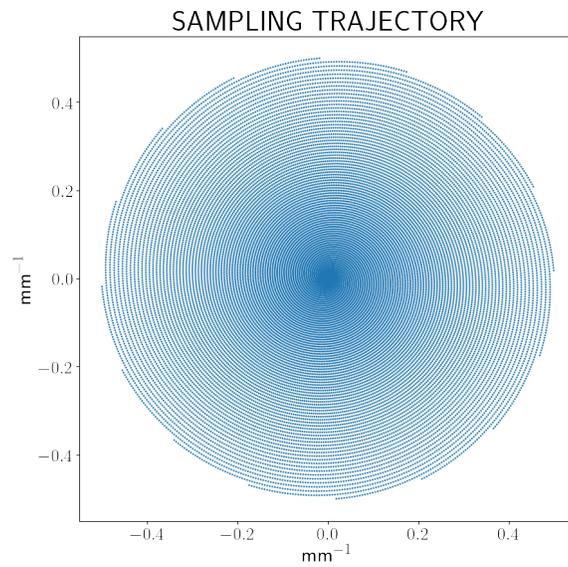


Figure 3-7: **Non-Cartesian MRI Trajectory.** This figure depicts the under-sampled trajectory used in (3.32).

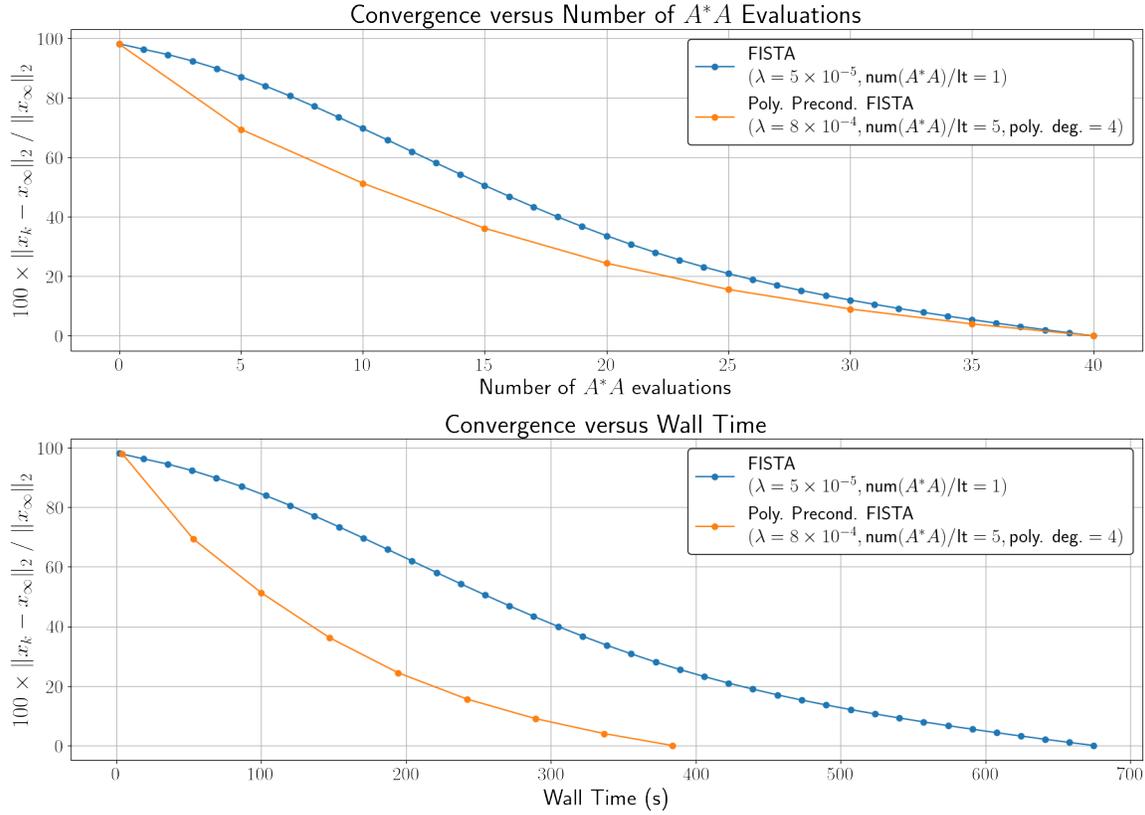


Figure 3-8: **Spatio-Temporal MRI Convergence Results.** This figure depicts the convergence results of evaluating (3.36a) for the first sub problem  $f_1$ . Given the chosen hyper-parameters, the error over iterations with respect to the last iteration of each respective method is plotted. The  $k^{th}$  iteration and last iteration are labelled as  $x_k$  and  $x_\infty$  respectively. The legend on the top right depicts the algorithm and the chosen hyper-parameters. The x-axis of the top and bottom subplots denotes the total number of  $A^*A$  evaluations and measured wall-times respectively. The circular-markings on each line denote the respective iteration points. The number of proximal operators evaluated by a point on the x-axis is equal to the number of iterations by that point.

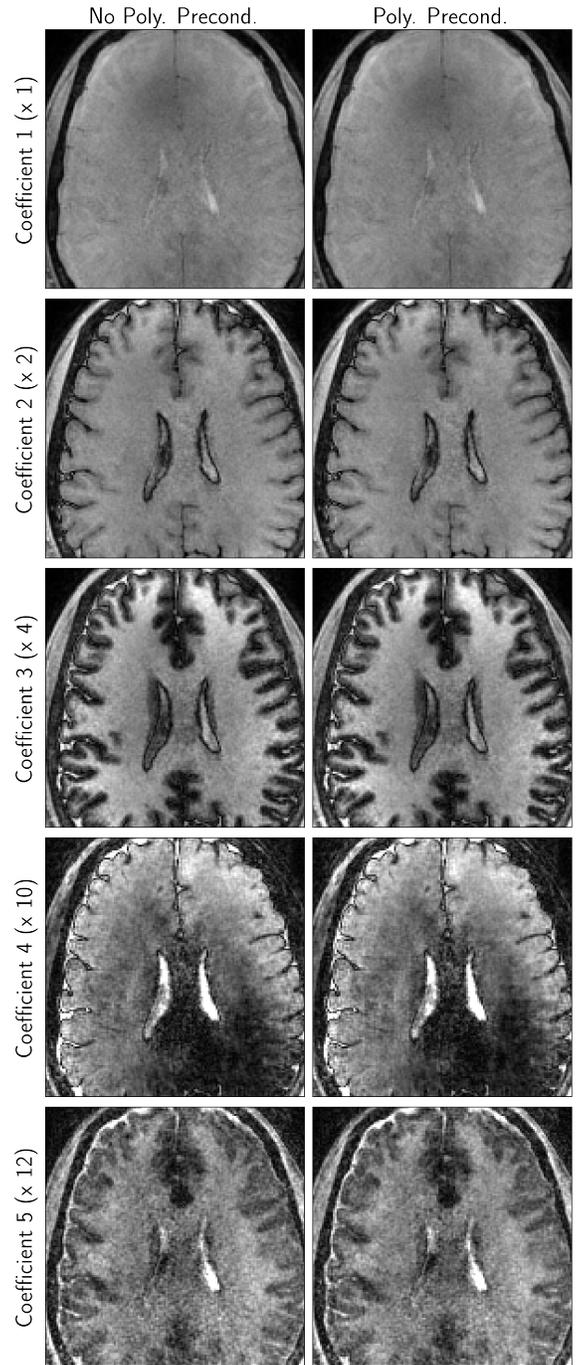


Figure 3-9: **Spatio-Temporal MRI Reconstruction Results.** This figure depicts the reconstructions after two ADMM iterations, where (3.36a) is evaluated using (top) FISTA and (bottom) Polynomial Preconditioned FISTA. The hyper-parameters for these results are depicted in figure (3-8).



## 4: Deep Learning Initialized Compressed Sensing

This chapter has Section (2.1) and Section (2.2) as “prerequisites”. Section (2.4) and Section (3.4.4) are recommended to have better context of the results.

### 4.1: Introduction

Recent developments in spatio-temporal MRI acquisition techniques have enabled whole-brain multi-parametric mapping in incredibly short exam times [94, 55, 61, 15, 102, 45, 111, 100, 5, 99, 48, 14]. These techniques leverage highly-efficient k-space encoding with subspace reconstruction (discussed in Section (2.4)) and carefully chosen regularization to achieve high-quality reconstruction without detrimental noise and artifact penalty despite high rates of acceleration [55, 94, 15, 102, 111, 99]. However, this comes at the cost of long reconstruction times for high-isotropic resolution volumetric imaging cases, making such methods difficult to integrate into clinical practice despite the high acquisition efficiency.

For example, consider the “Tiny Golden Angle Shuffling Spiral Projection Imaging Magnetic Resonance Fingerprinting” (TGAS-SPI-MRF[61, 15]) volumetric acquisition, which is the target application of this chapter. This application uses an optimized trajectory to achieve whole-brain multi-parametric mapping at 1 mm isotropic resolution in 1 minute of acquisition time using a LLR [94] regularized reconstruction implemented in BART [97]. With the settings of BART used in [15], the reconstruction takes around two hours on a CPU with around 130 GB of peak RAM usage, with the large dimensionality of the problem hindering out-of-the-box GPU utilization. This computational performance is achieved after BART’s default optimizations to improve reconstruction speed, such as FISTA method [4] for solving the LLR regularized optimization, as well as the combination of the “Toeplitz Point Spread Function” [101, 35, 3] and the “spatio-temporal kernel” [94] to reduce the per-iteration compute time.

Section (3.4.4) represents the first attempt at improving the reconstruction speed of this application, and this chapter explains the introduction of Density Compensation into (3.34). Assuming a parallel implementation and access to  $2 \times 20$  GB GPUs, the polynomial preconditioner enables a less than 15 minute reconstruction ( $8\times$  improvement) at higher quality compared to prior work [15]. However, this reconstruction performance is achieved by leveraging multiple high-end GPUs in parallel with minimal memory transfer between devices. Typical clinical workstations do not have access to such compute hardware, such as the  $2 \times 20$  GB GPUs used in Section (3.4.4). The wide distribution of this TGAS-SPI-MRF application requires a compute efficient reconstruction with minimal hardware requirements. Given the significant optimizations already performed to improve the iterative reconstruction speed, reaching this benchmark warranted the exploration of other solutions.

Concurrently with the development of fast spatio-temporal MRI acquisition, great progress has been made in utilizing deep learning for image reconstruction for acquisitions with high rates of acceleration [1, 41, 52, 88, 103, 42]. These methods commonly leverage an “unrolled” deep learning architecture, where the algorithm alternates between performing network inference and enforcing a physics-based Data-Consistency (DC) step akin to traditional iterative methods for solving regularized linear inverse problems [9, 59, 60, 58, 62, 34]. These works fall under the “physics-driven” classification proposed by [42], and will be referred to as such in the sequel. The integration of the DC term into these unrolled physics-driven methods has enabled the robust application of Deep Learning based MRI reconstruction without access to the copious amounts of training data typically required by deep learning methods [1]. These methods

have achieved excellent reconstruction performance at significantly faster processing times compared to their iterative convex algorithm counter-parts, which makes such frameworks a promising means of achieving fast spatio-temporal MRI reconstruction.

However, utilizing such unrolled physics-driven methods out of the box can be challenging depending on the dimensionality of the problem of interest. To use the target TGAS-SPI-MRF application as an example, the underlying desired signal is of dimensions  $256 \times 256 \times 256 \times 5$ , where  $256 \times 256 \times 256$  is the matrix size of the acquisition, and 5 denotes the number of subspace coefficients (that are needed to represent the temporal characteristics of the underlying signal) [55, 99, 111, 94, 102, 15]. This, along with using multiple receive channels and Non-Uniform Fourier Transform (NUFFT) operator (due to the Non-Cartesian trajectory utilized in TGAS-SPI-MRF[84, 61, 15]) dramatically increases the dimensionality of the reconstruction problem, making such unrolled physics-driven reconstructions extremely computationally challenging to train and deploy.

To give a concrete example, the current infrastructure at the Lucile Packard Children’s Hospital (Stanford, CA, USA) has multiple NVIDIA (R) GPUs with memories in the range from 6 GBs to 12 GBs. Preliminary exploration of leveraging multiple GPUs in parallel for image reconstruction were futile due to the synchronization costs (likely due to the large problem size of the application), and the 12 GB GPUs by themselves were not sufficiently large enough to utilize the unrolled physics-driven methods for the target application when utilizing a Residual Network[44] with 3D convolutions to perform model inference.

With these constraints in mind, this work proposes Deli-CS (pronounced “delicious”), which stands for “Deep Learning Initialized Compressed Sensing”. The goal of this framework is the rapid and highly compute-efficient subspace reconstruction of spatio-temporal MRI acquisitions (such as TGAS-SPI-MRF) with the goal of clinical deployment. This framework targets less than 6 GB peak GPU memory usage and approximately 10 minute reconstruction of the 1 minute TGAS-SPI-MRF acquisition that is comparable in quality to iterative LLR reconstruction of the same data.

While this work aims to describe a general deep learning initialization approach, prior work [37] has been proposed for rapid whole brain MRF reconstruction. Similarly to [16], [37] leverages an MRI trajectory that is not incoherent over the TRs, allowing a sliding window reconstruction to unalias the image. [37] improves over [16] by first projecting the signal onto a low-dimensional subspace to reduce the number of NUFFT required enabling faster reconstruction. However, as shown in [15], integrating coil sensitivity and subspace into the forward model (instead of a single coil combination and subspace projection in [37]) with incoherent sampling across TRs enables higher rates of acceleration with good reconstruction at just 1 minute of acquisition instead of 4 minutes and 55 seconds. This integration of coil information and subspace into the reconstruction enables better reconstruction of accelerated data at the cost of increased computation time. The proposed Deli-CS frameworks aims to fully leverage MRI physics and differs from [37] in the following key ways. First, it utilizes subspace and coil information to enable robust reconstruction at high accelerations. Second, the deep learning prediction is used to initialize an iterative reconstruction instead of being used as the output of the framework, which is expected to guard against potential hallucinations.

Given the strict 6GB GPU constraint of the proposed framework, it is worthwhile to discuss the motivations for enforcing the same. By being GPU efficient, the proposed framework is expected to scale well to ultra-high resolution sub-millimeter applications such as the 0.66 mm isotropic resolution “vista-MRF” acquisition for high-fidelity whole-brain myelin-water fraction (MWF) and  $T_1$ ,  $T_2$  and proton density mapping on a clinical 3T scanner [57]. Additionally, enforcing a minimal amount of GPU memory for training simplifies the process of “continuous training” to update the learned model to account for potential distribution shifts, which is an important factor to consider when deploying deep learning methods[30, 31, 26, 25, 27].

This chapter is organized as follows: First, the traditional LLR regularized reconstruction is implemented in Python using SigPy[74] in a GPU-efficient manner. This reconstruction is then augmented with Density Compensation to achieve an approximately 30 minute iterative reconstruction, which will be used as the baseline for comparison. Next, the Deli-CS framework is described, where a fast initial CG-based reconstruction is fed into a neural network that attempts to predict the final reconstruction. The training and inference are performed in a block-wise manner to significantly lower memory usage. Finally, since that the model does not have an integrated DC term and is not unrolled (this would be classified as “data-driven” in [42]), a “compressed sensing certification” step is proposed.

## 4.2: Background

The forward model of the TGAS-SPI-MRF application is similar what was discussed in (2.24). Let the subscript  $t$  denote the data acquired at the  $t^{\text{th}}$  TR (or time point), with  $T$  being the total number of TRs, and let  $K$  denote the number of coefficient images, or the rank of the low-rank subspace utilized. Then, the acquisition operator  $A$  is as follows:

$$A = \begin{bmatrix} \mathcal{F}_1 \\ \mathcal{F}_2 \\ \vdots \\ \mathcal{F}_T \end{bmatrix} S\Phi \quad (4.1)$$

Here,  $\mathcal{F}_t$  denotes the forward NUFFT operator at the  $t^{\text{th}}$  TR,  $S$  denotes the SENSE[85, 96] model and  $\Phi$  denotes the prior low-rank linear subspace with dimensions  $[T \times K]$ . The low-rank subspace  $\Phi$  is estimated by taking the Singular Value Decomposition (SVD) of signal dictionary generated from Bloch-simulations using the following realistic brain tissue parameters that matches prior work [15].

$$\begin{aligned} T_1 &\in \{20, 40, 60, \dots, 3000\} \cup \{3200, 3400, 3600, \dots, 500\} \\ T_2 &\in \{10, 12, 14, \dots, 200\} \cup \{220, 240, 260, \dots, 1000\} \\ &\quad \cup \{1050, 1100, 1150, \dots, 2000\} \cup \{2100, 2200, 2400, \dots, 4000\} \end{aligned}$$

The forward operation  $\Phi x$  recovers the TR images of the TGAS-SPI-MRF acquisition. A rank  $K = 5$  subspace was deemed sufficient in capturing the signal variation as per [15]. Please see [15] for more information about the acquisition and subspace forward model formulation. A modification made to the acquisition compared to the prior work [15] is the inclusion of a water-exciting rectangular pulse with duration of 2.38ms to minimize the excitation of the fat signal [72].

The linear inverse problem used to solve the subspace reconstruction is as follows:

$$\frac{1}{2} \|Ax - b\|_2^2 + \lambda \text{LLR}(x) \quad (4.2)$$

Here,  $\lambda$  is the regularization value and LLR denotes the locally low-rank constraint [94, 15].

[15] solved (4.2) with FISTA [4] using BART [97], which combines the Toeplitz Point Spread Function [101, 35, 3] operator with the subspace basis  $\Phi$  to derive a spatio-temporal operator that is similar to the spatio-temporal kernel in  $T_2$ -Shuffling [94] for faster evaluation of the normal operator ( $A^*A$ ) to reduce the per iteration compute time. While effective in reducing reconstruction times, utilizing the spatio-temporal kernel requires approximately 130 GB of CPU memory, increasing the computational demand of the reconstruction. To reduce the memory requirement, this work does not utilize the spatio-temporal kernel and instead relies on GPU processing for calculating the NUFFT using the default implementation in SigPy [74], and is further discussed below. For the 1 minute TGAS-SPI-MRF application considered in this work, this achieves similar iterations per second as the reconstruction in [15] while using less overall memory.

## 4.3: Methods

This work uses the Python-based SigPy [74] framework for the following experiments for its ease-of-use in prototyping the reconstruction. That being said, the methods proposed in this work still outperform the parameters used by the prior work [15] despite Python being a slower language than C. That being said, the planned future work is a C-based BART implementation of the methods discussed in this work, which is expected to provide additional speed-up due to C being a more compute-efficient language.

The sequel will first describe how the traditional LLR regularized reconstruction was optimized to improve iterative convergence, as this is leveraged by the Deli-CS framework in the ‘‘compressed sensing certification’’ step.

### 4.3.1 Memory Optimization

To reduce the memory requirements of  $A$ , the following standard optimizations are made. While not novel in itself, this subsection is presented for reproducibility's and posterity's sake as the following is essential to achieve low GPU memory usage.

First, the forward model  $A$  is split into smaller blocks so that each block can be iteratively and independently efficiently evaluated on the GPU. Second, to avoid expanding into the range space of  $\Phi$  ( $T \gg K$ ), the commutativity of the subspace operator  $\Phi$  and the NUFFT is used akin to the spatio-temporal kernel leveraged in  $T_2$ -Shuffling[94]. These two optimizations together result in the following forward model:

$$A(x) = \sum_{k,c} \underbrace{\Phi_k \mathcal{F}_{(1+2+\dots+T)} S_c}_{A_{(k,c)}} x_k \quad (4.3)$$

Here,  $x_k$  and  $\Phi_k$  denote the  $k^{th}$  coefficient image and the  $k^{th}$  column of  $\Phi$  respectively, and  $S_c$  denotes the  $c^{th}$  SENSE[85, 96] coil-sensitivity map. The individual smaller block linear operators,  $A_{(k,c)}$ , are evaluated one-by-one. The input to each block linear operator,  $x_k$ , is first transferred to GPU memory before applying the operator, and the resulting output,  $A_{(k,c)}x_k$ , is then transferred into CPU memory before the sum over coils ( $c$ ) and coefficients ( $k$ ) is applied.

The modified linear operator proposed in (4.3) achieves a computational speed of approximately 45 seconds per iteration of FISTA[4] when solving (4.2) for 1 minute TGAS-SPI-MRF acquisition, which coincidentally matches the seconds-per-iteration of FISTA achieved by BART on the CPU, while only utilizing approximately 16 GB of peak CPU memory and 4.5GB of peak GPU memory compared to 130 GB peak CPU memory. This was observed on a Linux workstation with an Intel (R) Xeon Gold 5320 and an NVIDIA (R) RTX A6000 GPU. Note that this performance is expected to vary between hardware, and that a BART implementation of the same will be faster still.

### 4.3.2 Density Compensation

Having achieved lower memory usage at comparable speed to the prior work[15], the next optimization targets improving the iterative convergence of FISTA. The acquisition operator  $A$  is ill-conditioned in that the difference between the largest and smallest eigenvalue of  $A^*A$  is large, yielding slow iterative convergence. Assuming 45 seconds per iteration and 200 over-all iterations, this results in approximately 2.5 hours required to reconstruct data. To improve this, first Pipe-Menon[79] density compensation was integrated into the optimization (4.2) as described in [3, 24, 23], yielding the following optimization:

$$\frac{1}{2} \|D^{1/2} (Ax - b)\|_2^2 + \lambda \text{LLR}(x) \quad (4.4)$$

Here,  $D$  is the Density Compensation array designed to target  $\mathcal{F}_{(1+2+\dots+T)}$  in (4.3) so that  $A^*DA$  has better conditioning. In the block linear operator form, this becomes:

$$A(x) = \sum_{k,c} D^{1/2} A_{(k,c)} x_k \quad (4.5)$$

Similarly to the prior section, the summing is performed in CPU memory with each block evaluated in GPU memory.

While the inclusion of  $D$  does in principle cause noise coloring, in practice, careful tuning of the LLR regularization value was found to provide suitable levels of denoising, resulting in high quality reconstructions in shorter times. This reflects the results discussed in [3], and demonstrates that Density Compensation can be leveraged to achieve high quality subspace reconstruction. The inclusion of density compensation into the optimization formulation is seen to significantly reduce the number of required iterations, with 40 iterations qualitatively determined to be sufficient for good reconstruction. In this work, the LLR block size used was 8.

Using Density Compensation significantly improves reconstruction quality compared to the prior work that used LLR [15] at much faster processing times. This is evidenced by Figure (4-1).

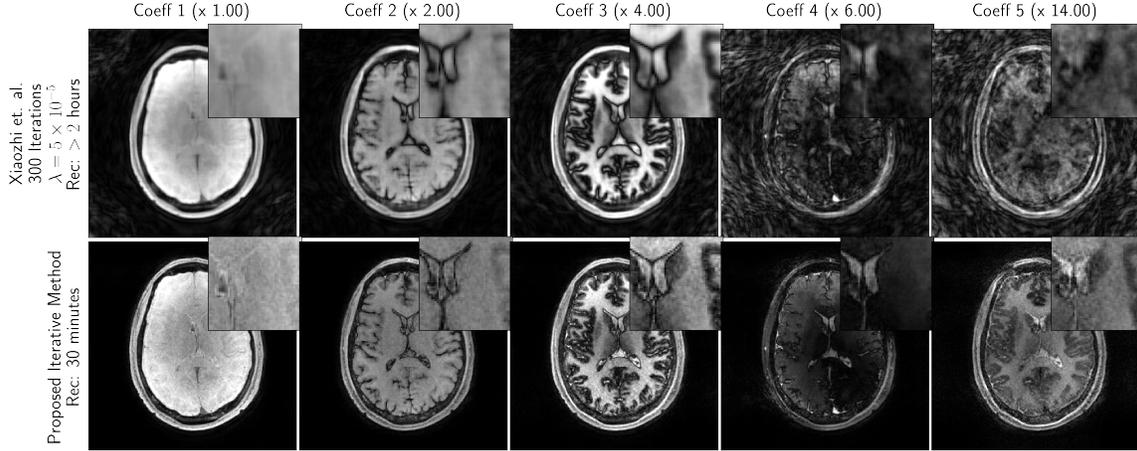


Figure 4-1: This figure compares the proposed reconstruction to the prior work [15]. Including Density Compensation (bottom row) into the linear inverse problem is seen to significantly improve the conditioning of the problem, resulting in much faster iterative convergence compared to not using Density Compensation (top row).

### 4.3.3 Field of View Processing

As discussed in [3], it is beneficial to reduce the matrix size of the reconstruction for lower memory usage and faster processing times. However, forcing a smaller field of view (FOV) without accommodating for the same is not advisable as signal from outside the FOV, particularly when the signal of interest is not properly centered (which happens in clinical practice), can result in artifacts during the reconstruction. For the TGAS-SPI-MRF brain imaging application, signal from the shoulders and neck are particularly troublesome. This is overcome by utilizing the automatic FOV shifting approach proposed in [3] that is augmented with a newly proposed coil compression method for additional robustness.

Prior to each TGAS-SPI-MRF experiment presented in Section (4.4), an unoptimized, 20 second, low resolution (6.9 mm isotropic) gradient echo (GRE) pre-scan with a large FOV of 440mm was performed. This pre-scan was used for automatic detection of the head position within the large FOV, so that the TGAS-SPI-MRF data could be automatically shifted to ensure that the brain, as well as the top of the head and nose, was within the smaller FOV used for the TGAS-SPI-MRF reconstruction. The automatic FOV centering was done by reconstructing the fully sampled, Cartesian, GRE image using a Fourier Transform, and performing a sum-of-squares combination of data from the multiple receive coils. The image was then flattened by taking a maximum intensity projection through the sagittal plane. The resulting 2D image was then smoothed, binarized, and a bounding box was calculated around the largest continuous area using the OpenCV toolbox [11]. A shift was then calculated to ensure the top and front of the bounding box was within the target field of view.

[3] derived the binary image from the eigenvalues of the estimated ESPIRiT maps, which is not used in this work as the target TGAS-SPI-MRF application utilizes 48 receive coils during the acquisition, making ESPIRiT computationally expensive. Additionally, the FOV shifting is performed *before* applying any coil compression techniques, which enables the following method for additional robustness.

Any signal originating from outside the bottom of the field of view after shifting (e.g. shoulders) was removed using region-optimized virtual coil compression [49] (ROVir) estimated from the GRE with the interference region set to any area outside the target FOV for the TGAS-SPI-MRF acquisition. The data was pre-whitened as described in [49] before ROVir was applied. Signal outside the field of view was removed by throwing away 8 virtual ROVir coils that contained the most signal in the interference region. After that, SVD compression to 10 virtual channels of the remaining channels were performed to reduce the size of the computation. The complete coil compression matrix containing coil whitening, ROVir, and SVD compression was calculated based on the GRE, and used for compression of the TGAS-SPI-MRF data.

The reconstruction, automatic FOV shifting, and coil processing matrix calculation for the GRE data

takes less than 30 seconds, and is run while the TGAS-SPI-MRF data is being acquired.

#### 4.3.4 Deli-CS

With Deli-CS, the goal is to achieve an approximately 10 minute reconstruction of TGAS-SPI-MRF data that is of comparable quality to the traditional reconstruction with minimal memory footprint. To achieve this, the following design pillars are utilized:

1. *Fully Leveraging MRI Physics.* As a first step, in order to utilize prior coil sensitivity and the subspace information as much as possible, an approximate reconstruction is run by setting  $\lambda = 0$  in (4.4). This ordinary least squares optimization uses Conjugate Gradient (CG), and allows the multiple receive channels to perform an initial resolving of the subspace coefficients through Parallel Imaging [85, 38, 96, 109]. CG was used as it was observed to be faster than the LLR reconstruction for the initial resolving of the data while still doing better with respect to the zero-filled image. By not using LLR regularization, the input to the next step suffers from temporal-aliasing artifacts and increased noise, which the following deep learning step is expected to robustly mitigate. The resulting CG reconstruction will be referred to as “Deli-CS Input”.
2. *Block-based Data-Driven Deep Learning.* A deep learning network is trained to denoise and dealias the input CG image. The model is both trained and deployed in a “block-wise” manner to reduce the memory and training data requirements, and is consequently *not* integrated with DC terms in an unrolled manner. This is done to avoid needing to perform back-propagation during training through the high-dimensional acquisition operator  $A$ , which is challenging to do even with a GPU with large memory capacity. This deep learning methodology is similar to the “data-driven” classification in [42], except that instead of taking in as input the adjoint image ( $A^* b$ ), which is commonly called the “zero-filled image”, the input is the above CG reconstruction which has reduced artifacts, particularly for the first three coefficients. The block-based processing proposed in this steps allows the deep learning model to be trained and deployed efficiently with under 6 GB of GPU memory. The resulting inference will be referred to as “Deli-CS Prediction”.
3. *Compressed Sensing Certification.* Since the above network is block-based and data-driven, the inferred reconstruction suffers is possibly susceptible to hallucinations. To protect against the same, the inferred result is used to initialize (4.4) solved with an iterative reconstruction. By initializing the iterative reconstruction with the inferred result, the number of iterations required to converge is significantly reduced. Additionally, the resulting image is “Compressed Sensing Certified” in the sense that the resulting image satisfies the same convergence criterion as the traditional reconstruction achieved when solving (4.4). This step will be referred to as the “refinement” step, with the resulting reconstruction denoted “Deli-CS Refined”.

In practice, block-wise network inference results in blocking artifacts, even when passing a large block into the network (say  $64 \times 64 \times 64$ ) and cropping the inferred result to a smaller inner block ( $32 \times 32 \times 32$ ) to account for boundary conditions. This process was further augmented with overlapping blocks assuming a stride of 16 to attempt to average out the artifacts, but was not successful. In particular, this instead overly smoothed the inference result to the point where the consequent refinement step required a non-trivial number of iterations to add the details back in. Using a larger block (say  $128 \times 128 \times 128$ ) was observed to degrade the quality of the network prediction while adding to the computational burden of inference. More importantly, in the above two cases, the iterative refinement step was not able to resolve said artifacts quickly enough for the initialization to be useful.

Instead, the following procedure is used. First, the network inference is performed using blocks of size  $(8 \times 8 \times 8)$  with strides  $(4 \times 4 \times 4)$ , which does still induce blocking artifacts. However, by utilizing a small stride value, these artifacts manifest as high frequency artifacts that are subdued by performing one round of wavelet thresholding at the cost of increased blurring, which is circumvented by the refinement iterations. This processing is verified by Figure (4-2).

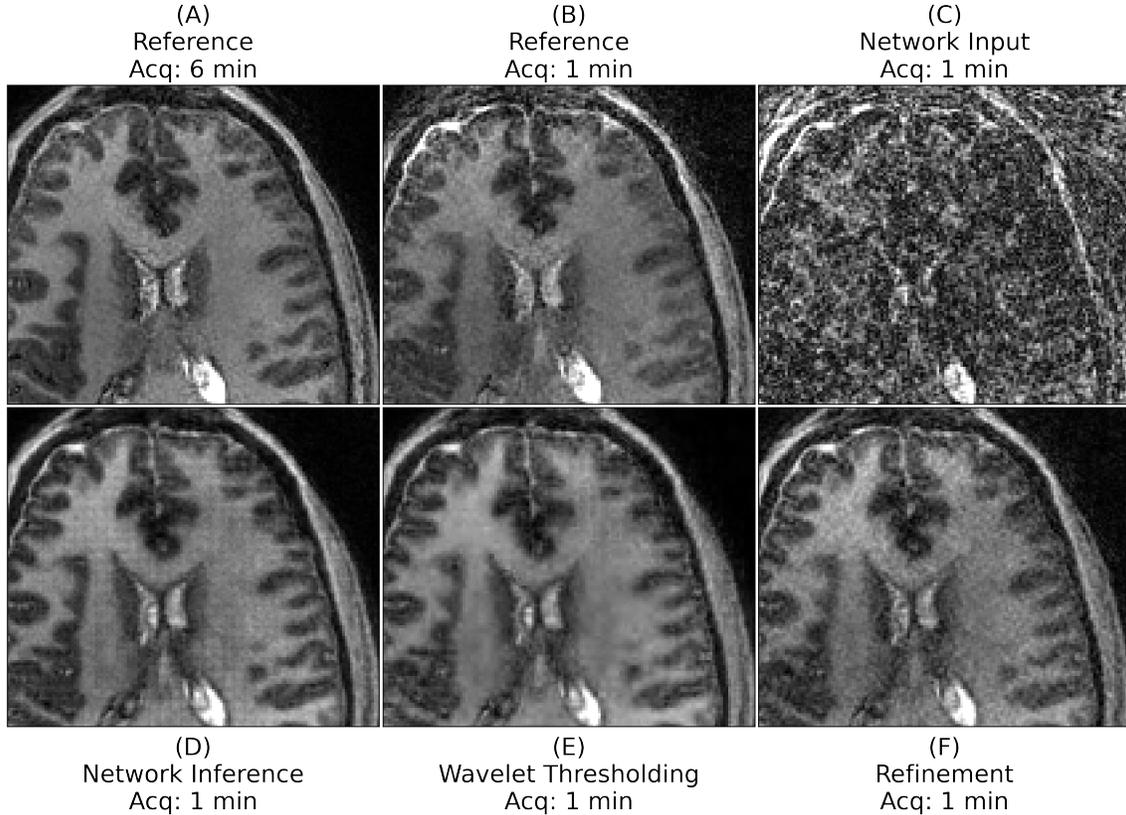


Figure 4-2: This figure depicts the third coefficient of various reconstruction to demonstrate how artifacts from block-wise network processing are alleviated. (A) is the reference “target” reconstruction from the 6 minute acquisition; (B) is the reference reconstruction from the 1 minute acquisition; (C) is the CG reconstruction that is input to the network; (D) is the prediction; (E) is the result after wavelet thresholding; and (F) is the result after 7 FISTA iterations of refinement. The input to the network was blocks of size  $(8 \times 8 \times 8)$  with strides  $(4 \times 4 \times 4)$ . (E) suffers from blocking artifacts that are suppressed by the wavelet thresholding (E). The resulting undesirable loss of detail is overcome by performing refinement iterations (F).

### 4.3.5 Basis Balancing

The subspace basis is estimated by performing a SVD of a dictionary of realistic signal evaluations generated using the Bloch equation and using singular vectors corresponding to the top five singular values to form  $\Phi$  [15]. This rank-truncation level was deemed sufficient in capturing the signal variation as per [15], and the parameters used to derive the dictionary was presented in Section (4.2). Using this basis directly for the initial CG reconstruction results in very low signal level in the fourth and fifth coefficients as most of the signal is already captured in the first three basis. This is shown in Figure (4-3)(A). To ensure each coefficient image contributes roughly equally to the objective function when model training and to more equally distribute the artifacts across coefficients, a “basis balancing” heuristic is proposed to equally distribute the energy across all the coefficients.

The low-rank basis  $\Phi$  of dimensions  $T \times K$  with  $\Phi_k$  denoting the  $k^{th}$  column of  $\Phi$ . Let  $B$  denote the new basis with  $b_k$  denoting the  $k^{th}$  column. In order to balance  $B$ , the columns  $b_k$  are derived so that  $b_k$  consists of equal contributions from each column of  $\Phi$ . This translates to an equality constraint on the magnitude of the inner products. That is to say, for some constant  $\alpha$ ,

$$|\langle b_p, \Phi_q \rangle| = \alpha \text{ for all } p, q \quad (4.6)$$

In other words,  $B^* \Phi$  is a matrix where the magnitude of each matrix is the same.

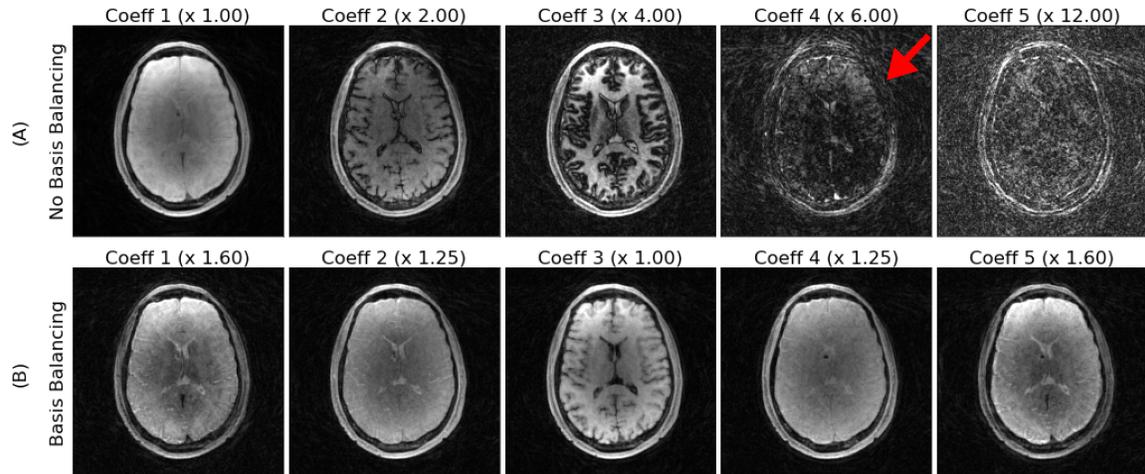


Figure 4-3: This figure depicts how the underlying images look after the initial least squares reconstruction (A) without basis balancing, and (B) with basis balancing. Note the skull signal dropout in coefficient 4, marked with a red arrow.

One such matrix that satisfies this property is the Discrete Fourier Transform (DFT) matrix. Let  $\Theta$  be the DFT matrix of dimensions  $K \times K$ . Then, the balanced basis  $B$  can be derived as:

$$B = \Phi \times \Theta \tag{4.7}$$

With  $\Theta$  normalized to have unitary columns,  $B$  is also an orthonormal matrix. Additionally, since the columns  $B$  are constructed from a linear combination of the columns of  $\Phi$ ,  $B$  and  $\Phi$  span the same subspace. In other words,  $B$  is a linear combination of the columns of  $\Phi$  and thus a reconstruction with or without basis balancing should give the same result up-to a change-of-basis transformation. Since  $B$  is derived from  $\Theta$ ,  $\Theta^{-1}$  can be used to perform a linear change of basis from  $B$  to  $\Phi$ . This was leveraged by the previous demonstrations as well as in (3.34), and will be used for rest of this chapter. In other words, all reconstructions presented in this chapter used the balanced basis  $B$ , but the coefficient images with respect to  $\Phi$  are shown after applying the change-of-basis operation  $\Theta^{-1}$  to the recovered coefficients. This is done for interpretability and ease-of-comparison against prior reported reconstructions in [15]. Using  $B$  for the initial CG reconstruction yields coefficient images of roughly equal signal level with no one coefficient suffering from significantly more artifacts compared to the others as shown in Figure (4-3)(B).

#### 4.4: Experiments

To train and verify the Deli-CS framework, data from 14 healthy volunteers were acquired on a 3T Premier MRI scanner (GE Healthcare, Waukesha, WI) and a 48-channel head receiver-coil. GRE and TGAS-SPI-MRF (with 500 TRs)[15] data were acquired. The TGAS-SPI-MRF acquisition time was 6-minutes, acquired resolution was 1 mm isotropic, and FOV was 220 mm isotropic. The TGAS-SPI-MRF acquisition parameters are described in Section (3.4.4). The data was retrospectively sub-sampled to simulate a 1-minute acquisition. The data were partitioned as 10 training, 2 validation and 2 testing subjects. Coil sensitivity maps were estimated with JSENSE[109]. The dictionary and subspace were generated as describe in Section (4.2), and the template matching was used to estimate the  $(T_1, T_2)$  parametric maps.

The reported  $\lambda$  values for all reconstructions are after right-hand-side of the DC term of (4.4) (i.e.  $D^{1/2}b$ ) is normalized to have unitary  $l_2$ -norm. The reference LLR reconstruction was performed on the 6-minute data with the LLR block size of 8 and a  $\lambda$  value of  $3 \times 10^{-5}$  with 40 FISTA iterations. Note that the estimated density compensation function  $D$  is different for the 6 minute and 1 minute data. The matrix size was  $256 \times 256 \times 256$ . This will be referred to as the “target” reconstruction. The 6-minute reconstruction was used as the label for network training instead of the 1-minute reconstruction as the

6-minute result has higher SNR and is less likely to have residual aliasing artifacts, especially for the low-energy coefficients. Future work is to use this 6-minute data to design a robust deep-learning unrolled method that improves upon the LLR reconstructions to hopefully achieve further reduction in acquisition time.

The retrospectively under-sampled 1-minute data LLR reconstruction was also performed for comparison using an LLR block size of 8 and a  $\lambda$  value of  $5 \times 10^{-5}$  with 40 FISTA iterations. The assumed matrix size was  $256 \times 256 \times 256$ .

For Deli-CS, the initial CG reconstruction was performed with 6 CG iterations, which was qualitatively determined to be suitable trade-off between noise amplification in the fourth and fifth coefficients (with respect to the SVD basis) and the initial resolving of the first three coefficients. This is shown in Figure (4-4). Corresponding blocks of dimensions  $64 \times 64 \times 64 \times 5$  are extracted from the CG reconstruction and the

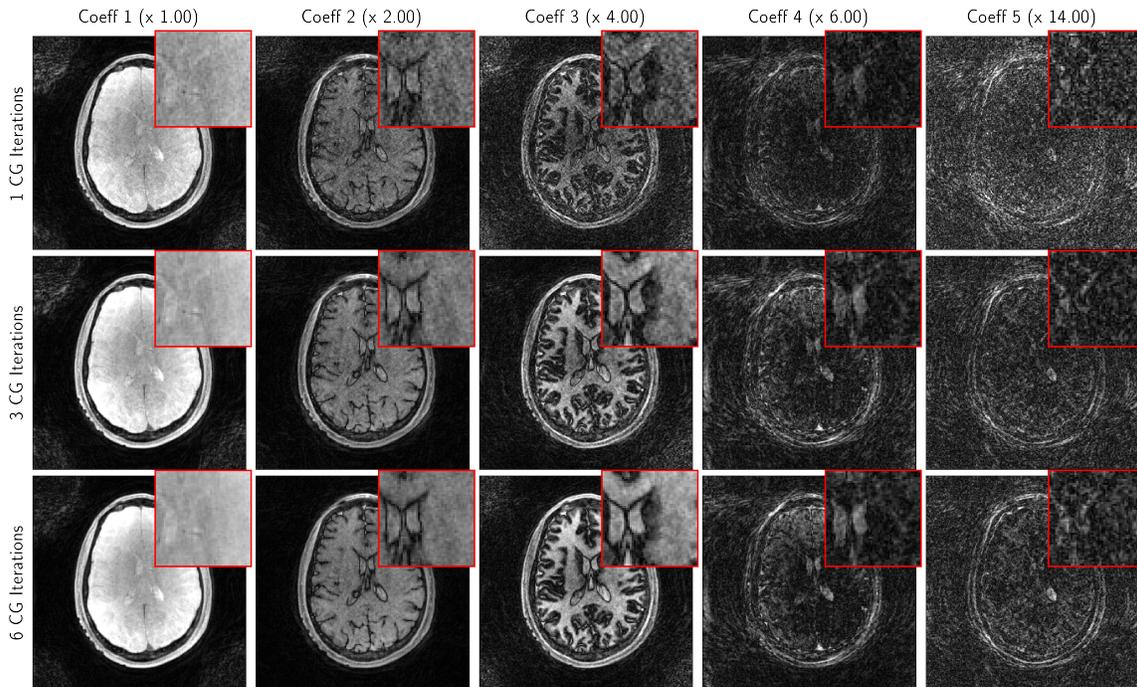


Figure 4-4: This figure depicts how the initial number of CG iterations were determined. Note that Coefficients 2 and 3 are well resolved at 6 iterations compare to 1 and 3. Higher iteration values increase computation time and may induce noise amplification in the latter coefficients.

target reconstruction. For data augmentation, random flips, transposes and shifts are performed. Each block is normalized so that the middle coefficient has a maximum value of 1. After normalization, if the middle coefficient's standard deviation is below 0.3, the block is discarded. This simple filter ensures that the network avoids learning from regions with no signal. The blocks are split into real and imaginary components, and concatenated along the subspace dimension. These blocks are piped into ResNet [44] with 3D convolutions, where the channel dimension corresponds to the subspace dimension. The ResNet utilized 2 residual blocks with a filter size of 3, a feature dimension of 128 and ReLU activation. The model was implemented in PyTorch Lightning [77, 33] and trained for 2000 epochs using the ADAM optimizer with a learning rate of  $1 \times 10^{-5}$  that decays by 0.5 every 1000 steps. The training utilized under 5 GB of GPU memory.

For the final step of Deli-CS, the model prediction is used to initialize (4.4) after performing wavelet thresholding with a regularization value of  $7.5 \times 10^{-5}$ . The iterative reconstruction used for refinement utilized the same parameters as the reference 1-minute reconstruction. Since the prediction is close to the final solution, seven iterations of refinement are deemed sufficient to achieve to similar reconstruction quality.

## 4.5: Results

The reconstruction of the first test subject is depicted in Figures (4-5) to (4-7). The reconstruction of the second test subject is depicted in Figures (4-8) to (4-10). The corresponding parametric fittings are shown in Figures (4-11) and (4-12) and Figures (4-13) and (4-14). A magnitude threshold was applied to the figures to zero-out regions outside the FOV of the signal for presentability.

On a Linux workstation with an Intel (R) Xeon Gold 5320 and an NVIDIA (R) RTX A6000 GPU, the full GRE coil compression FOV processing took approximately 14.5 seconds; the JSENSE estimation took approximately 21.5 seconds; the reference iterative reconstruction took 27 minutes; the Deli-CS initial CG reconstruction took approximately 3 minutes and 10 seconds; the Deli-CS model inference with wavelet thresholding took approximately 1 minute and 25 seconds; the final refinement step with seven iterations took approximately 5 minutes and 6 seconds.

The whole pipeline takes approximately 30.4 minutes using the reference iterative reconstruction, and approximately 10.3 minutes using Deli-CS. Thus, the initialization approach enables 3x faster processing times.

As evidenced by Figure (4-12), the refinement step adds missing features that were over smoothed from the network prediction and wavelet thresholding process.

## 4.6: Discussion

This work presented a framework for MRI reconstruction that targets high dimensional applications like volumetric non-Cartesian spatio-temporal subspace reconstruction, with the goal of reconstructing said application in an approximately 10-minute time frame with modest hardware requirements. This was achieved with a block-based deep learning initialization approach, where the deep learning prediction was used to jump-start a regularized linear inverse problem.

In order to achieve the less than 6 GB GPU condition, it is necessary to apply network inference in a block wise manner which unfortunately introduces blocking artifacts. In order to overcome this, this work leverages block-wise inference using small block sizes which resulted in “high-frequency” artifacts that are well subdued by performing wavelet thresholding. This is at the cost of introducing some blurring, which is accounted for by the refinement step. The number of iterations of refinement is a hyper-parameter that determines the trade-off between reconstruction speed and image sharpness. The seven iteration refinement outperforms the prior reconstruction [15] as evidences by Figure (4-1), but is a bit more over-smoothed when looking at the parametric  $T_2$  compared to the iterative reconstruction with Density Compensation. Adding more iterations will alleviate this issue at the cost of more reconstruction time. This is because, in principle, the optimization in (4.4) has a unique solution that the iterations will converge to, regardless of initialization. However, the reconstruction of the fourth and fifth coefficient images are better than the reference 1 minute case as it is still benefits from the prediction power of the deep learning component.

The basis balancing approach proposed is seen to spread the energy across the coefficient images, avoiding excessive aliasing artifacts.

With respect to the fitted parametric maps, there are no significant differences between the reference 10 minute reconstruction, the reference 30 minute reconstruction and the Deli-CS refinement results. This is likely because most of the information for standard quantitative map fitting of  $T_1$  and  $T_2$  maps exists in the first three coefficients that contain the majority of the energy. However, the good recovery of the fourth and fifth coefficients is expected to improve more advanced quantitative parameter fitting, such as in multicomponent modeling and multidimensional correlation spectroscopic imaging [50]. In particular, since a voxel typically consists of multiple tissue types, the better resolved fourth and fifth coefficients are expected to allow for the better fitting of multiple  $T_1$  and  $T_2$  values per voxel.

This chapter did not leverage the polynomial preconditioner as discussed in Chapter 3 due to the following reasons. First,  $A^*DA$  has a narrow eigenvalue spectrum which reduces the theoretical efficacy of the preconditioner. Next, the memory-efficient implementation of  $A$  resulted in  $A^*DA$  taking longer to evaluate compared to the LLR proximal operator. With these two factors, the polynomial preconditioner imparted a minor  $\sim 1.5\times$  reduction to the iterative reconstruction times. A more nuanced reason is that an iterative LLR reconstruction typically leverages some kind of spatial shifting over iterations to

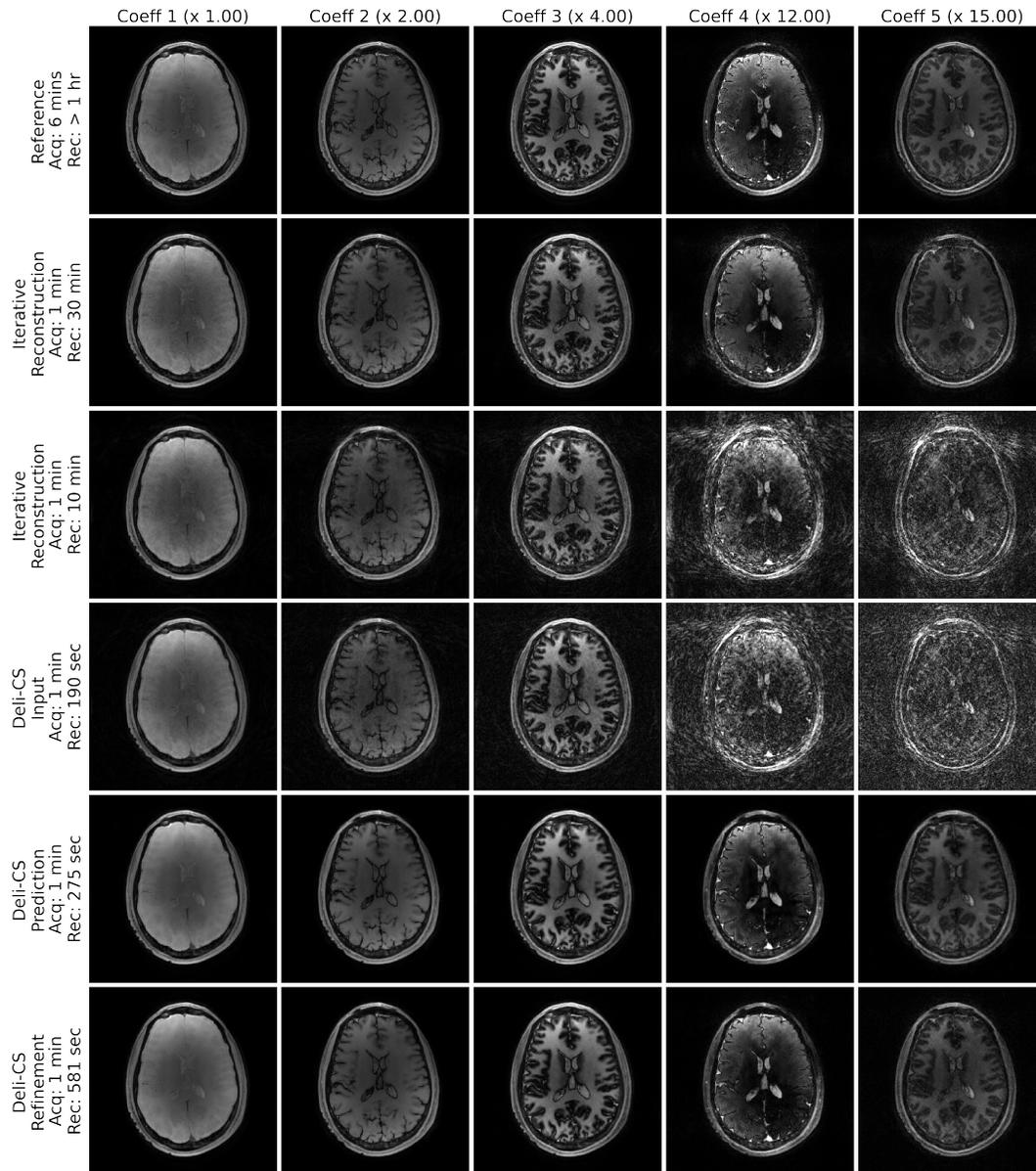


Figure 4-5: This figure compares the coefficient axial images recovered from the TGAS-SPI-MRF data using various methods for the first test subject. The first row denotes the reference LLR reconstruction of the 6-minute data acquisition, and the second row denotes the LLR reconstruction of the retrospectively under-sampled 1-minute acquisition. The third row depicts the LLR reconstruction after 10 minutes of processing for comparison. The remaining rows depicts the various steps of Deli-CS and the time taken to reach that point. The fourth row shows the initial CG reconstruction, the fifth row shows the model inference, and the ultimate row shows the reconstruction after iterative refinement. “Acq” denotes the acquisition time, and “Rec” denotes the reconstruction (or inference, for Deli-CS Prediction) time. Note that the Deli-CS timings reported include the previous steps.

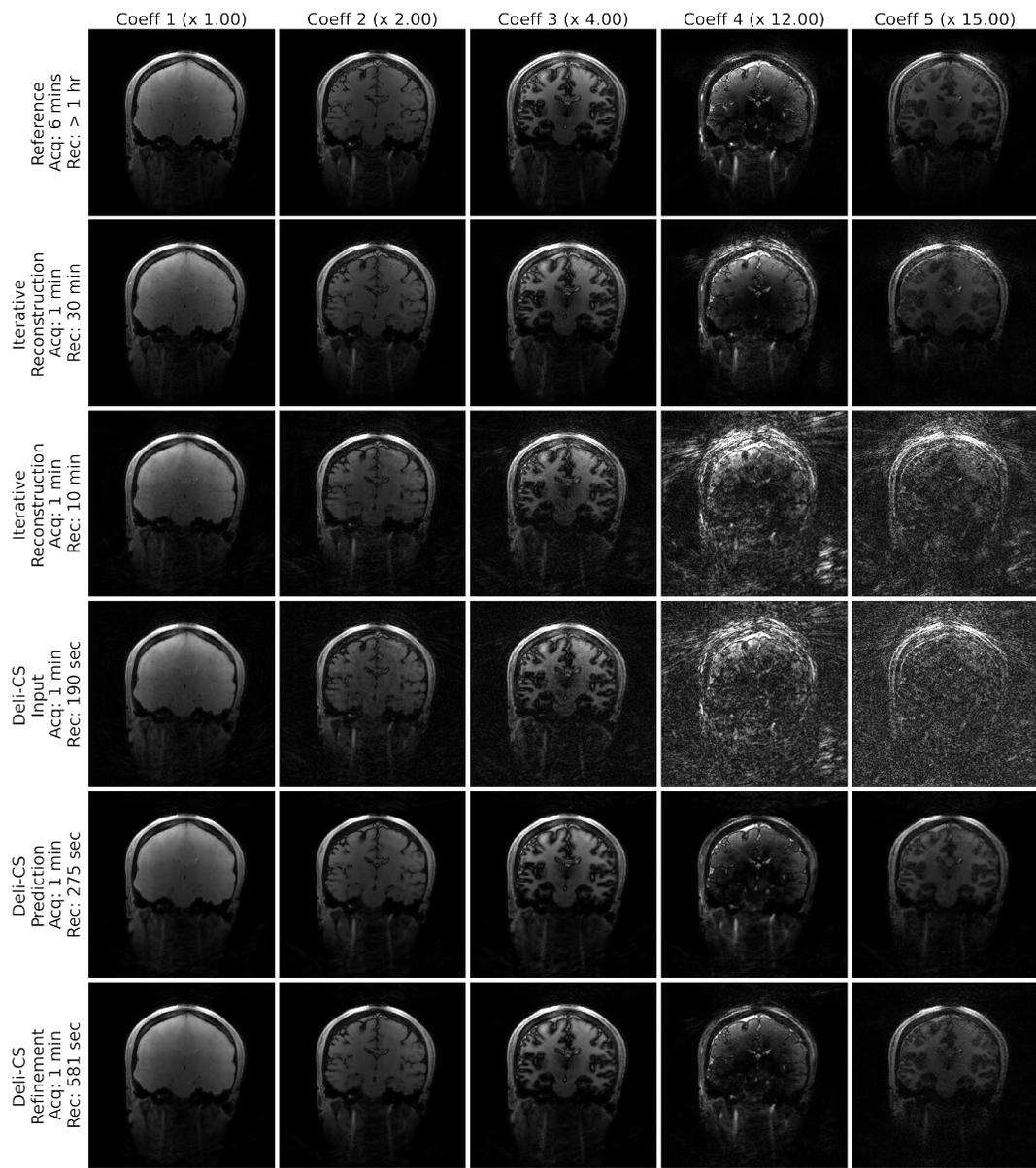


Figure 4-6: This shows the coronal slices of the reconstructions presented in Figure (4-5).

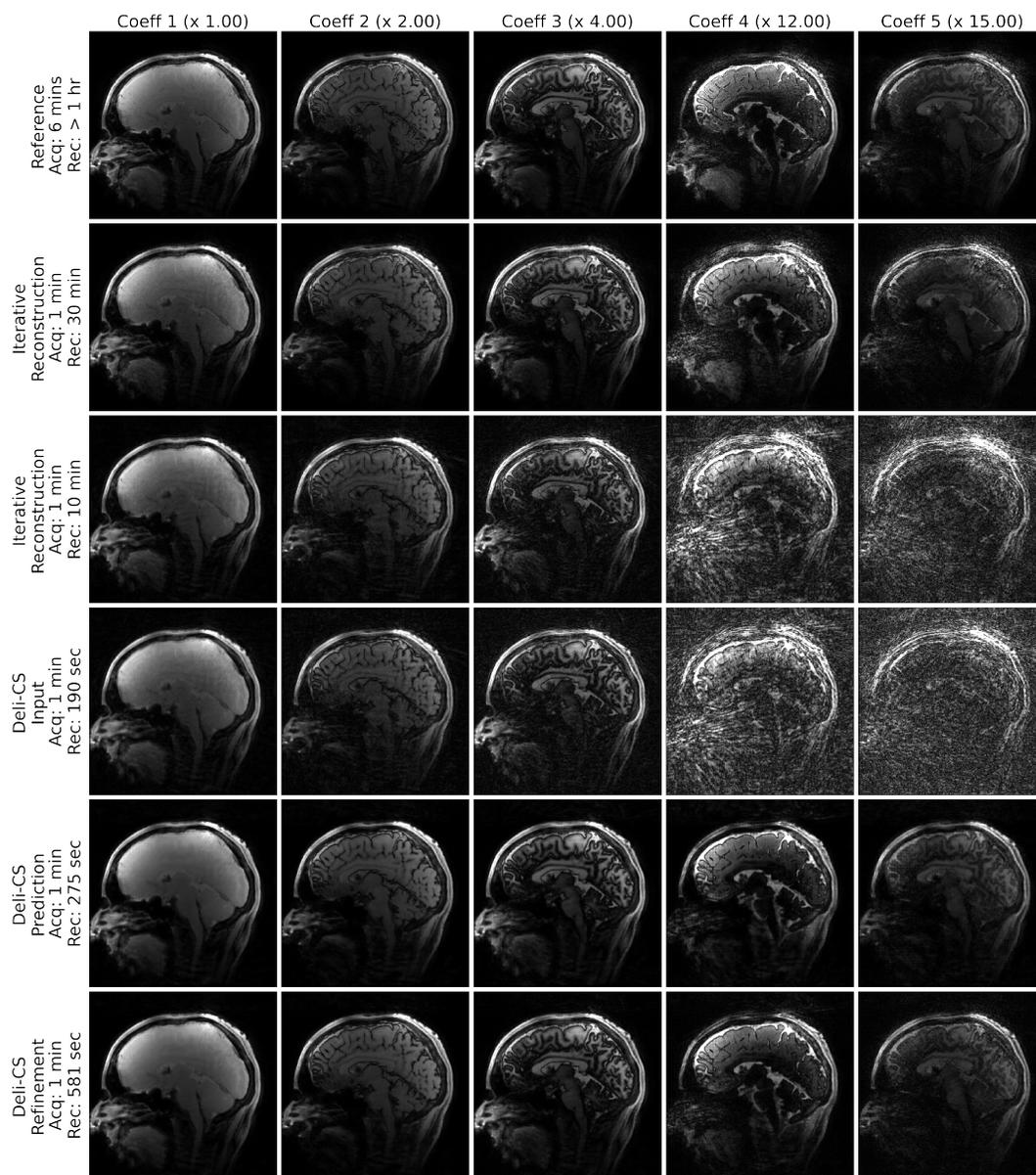


Figure 4-7: This shows the sagittal slices of the reconstructions presented in Figure (4-5).

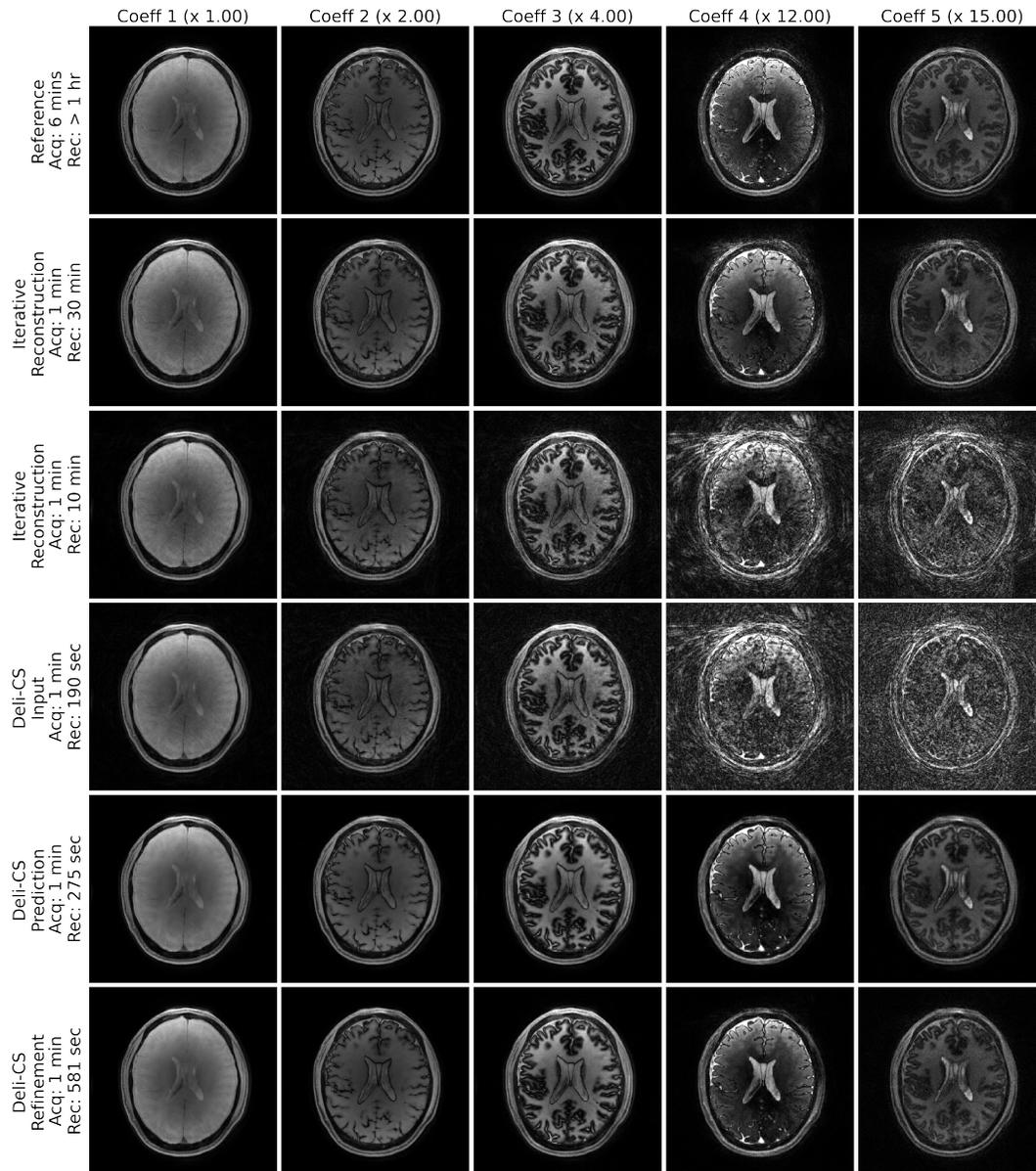


Figure 4-8: This figure compares the coefficient axial images recovered from the TGAS-SPI-MRF data using various methods for the second test subject. The first row denotes the reference LLR reconstruction of the 6-minute data acquisition, and the second row denotes the LLR reconstruction of the retrospectively under-sampled 1-minute acquisition. The third row depicts the LLR reconstruction after 10 minutes of processing for comparison. The remaining rows depicts the various steps of Deli-CS and the time taken to reach that point. The fourth row shows the initial CG reconstruction, the fifth row shows the model inference, and the ultimate row shows the reconstruction after iterative refinement. “Acq” denotes the acquisition time, and “Rec” denotes the reconstruction (or inference, for Deli-CS Prediction) time. Note that the Deli-CS timings reported include the previous steps.

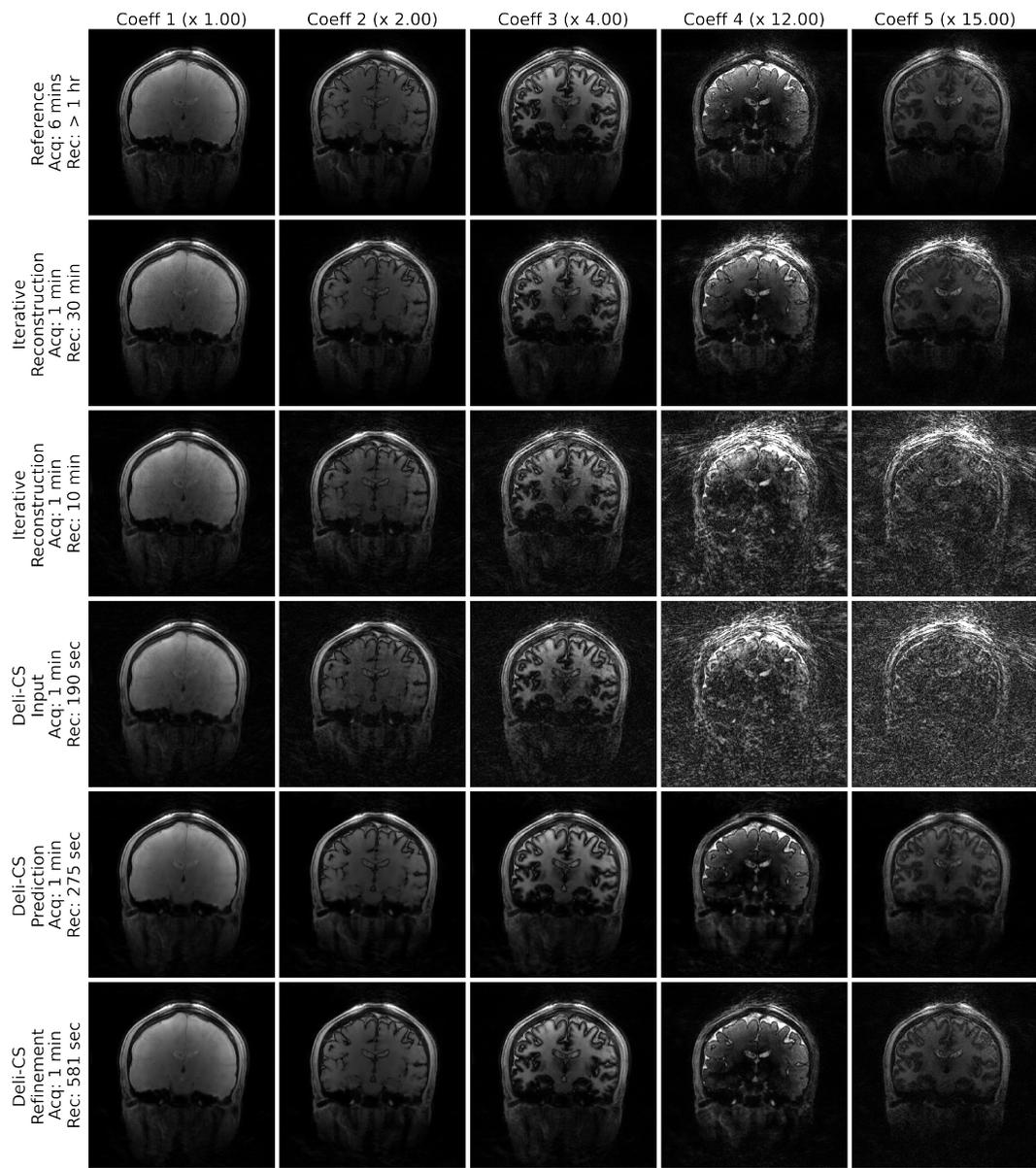


Figure 4-9: This shows the coronal slices of the reconstructions presented in Figure (4-8).

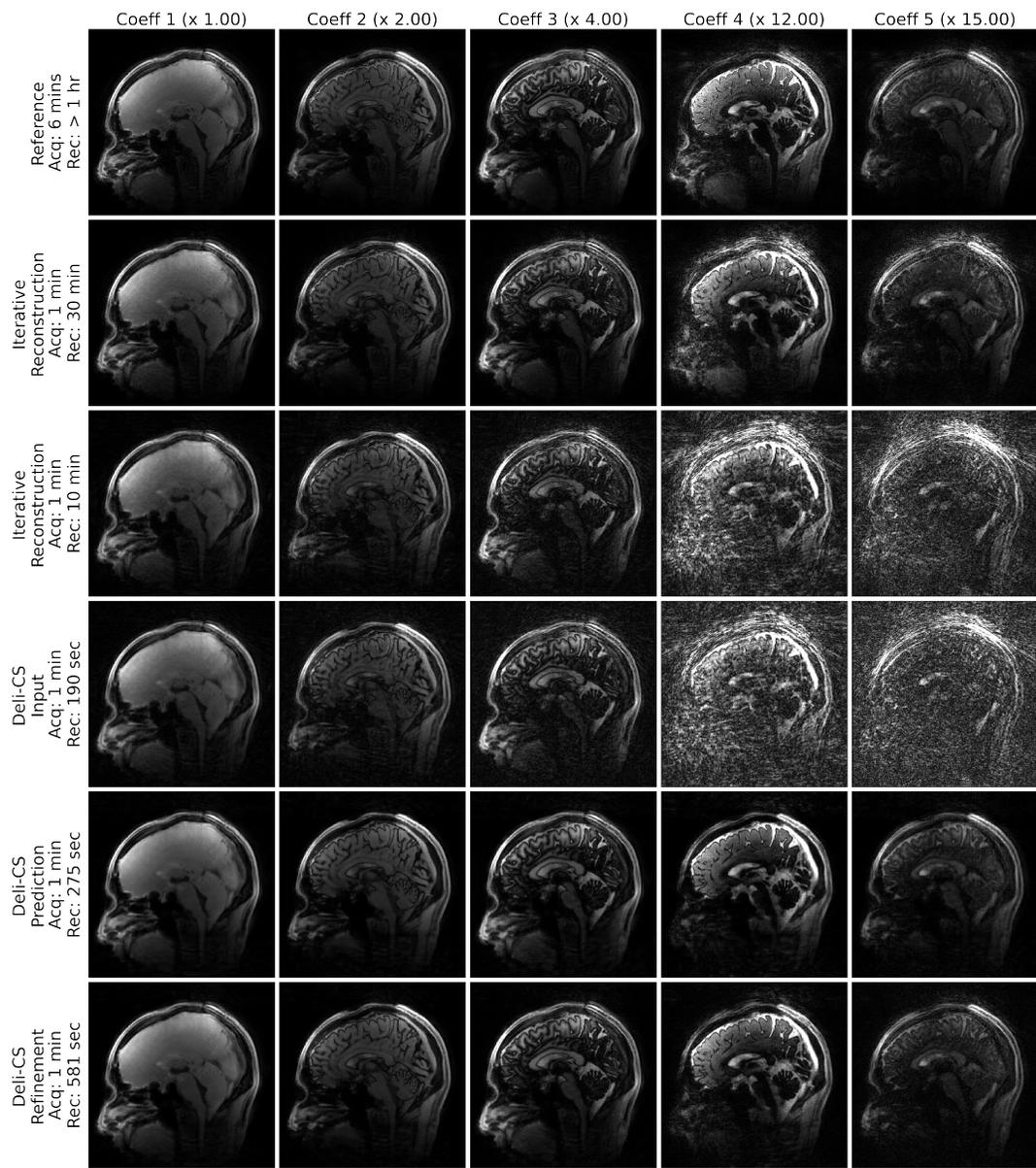


Figure 4-10: This shows the sagittal slices of the reconstructions presented in Figure (4-8).

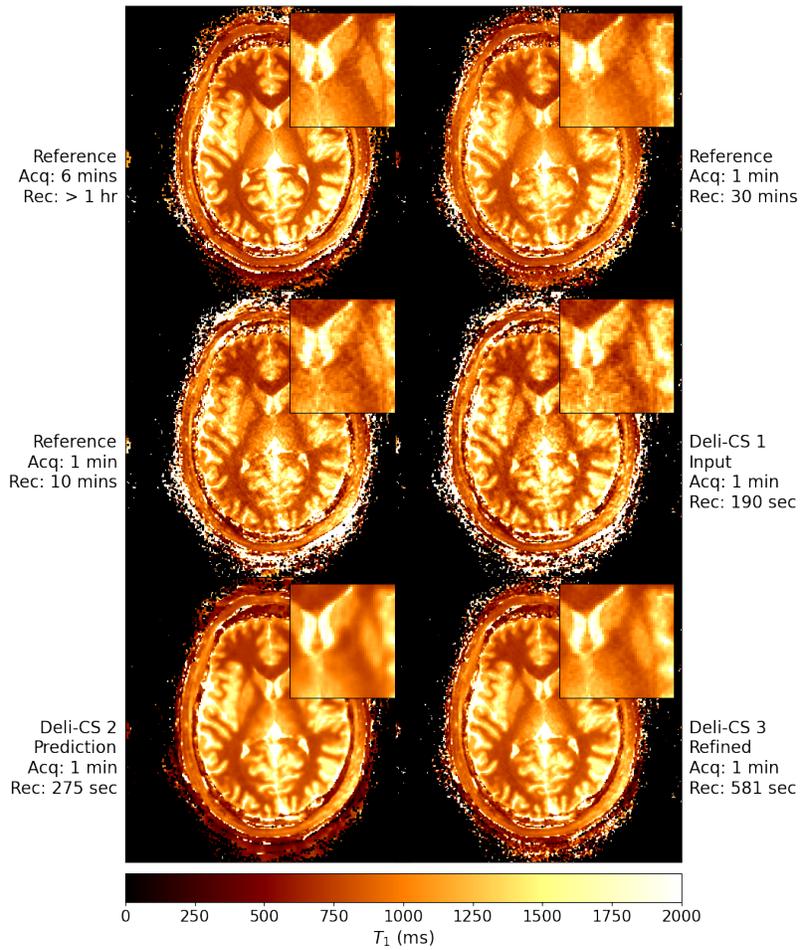


Figure 4-11: This figure compares the  $T_1$  parameter fittings of TGAS-SPI-MRF reconstructions depicted in Figure (4-5). “Acq” denotes the acquisition time, and “Rec” denotes the reconstruction (or inference, for Deli-CS Prediction) time. The zoomed in region demonstrates how the refinement step helps recover sharp features that may be over-smoothed by the Deli-CS Prediction step. Note that the Deli-CS timings reported include the previous steps.

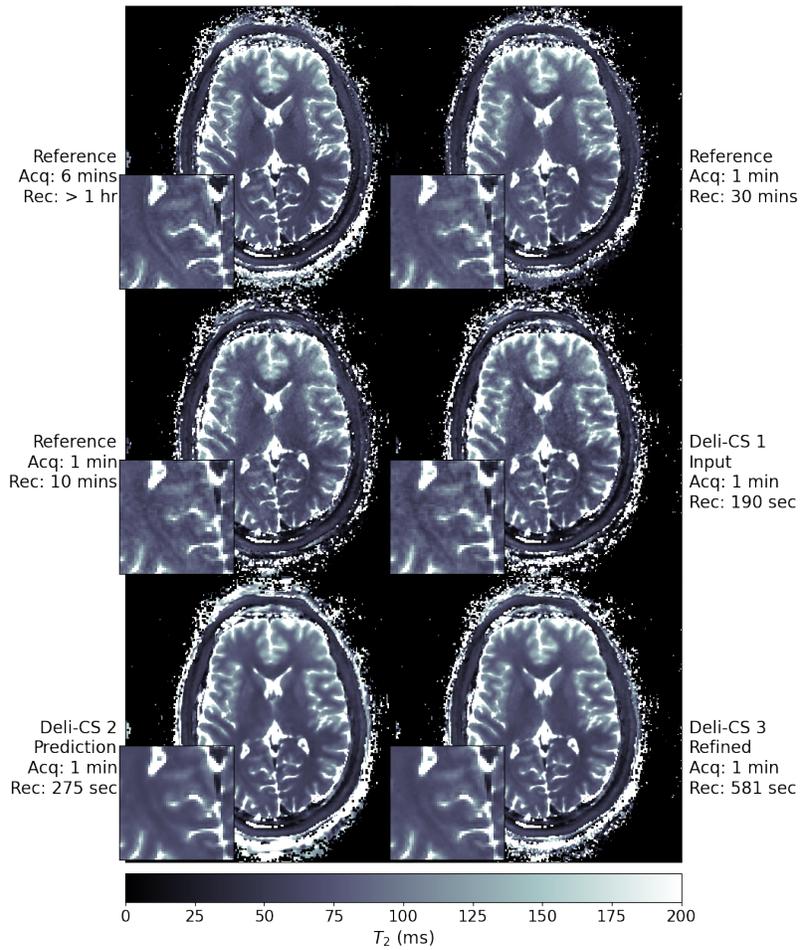


Figure 4-12: This figure compares the  $T_2$  parameter fittings of TGAS-SPI-MRF reconstructions depicted in Figure (4-5). “Acq” denotes the acquisition time, and “Rec” denotes the reconstruction (or inference, for Deli-CS Prediction) time. The zoomed in region demonstrates how the refinement step helps recover sharp features that may be over-smoothed by the Deli-CS Prediction step. Note that the Deli-CS timings reported include the previous steps.

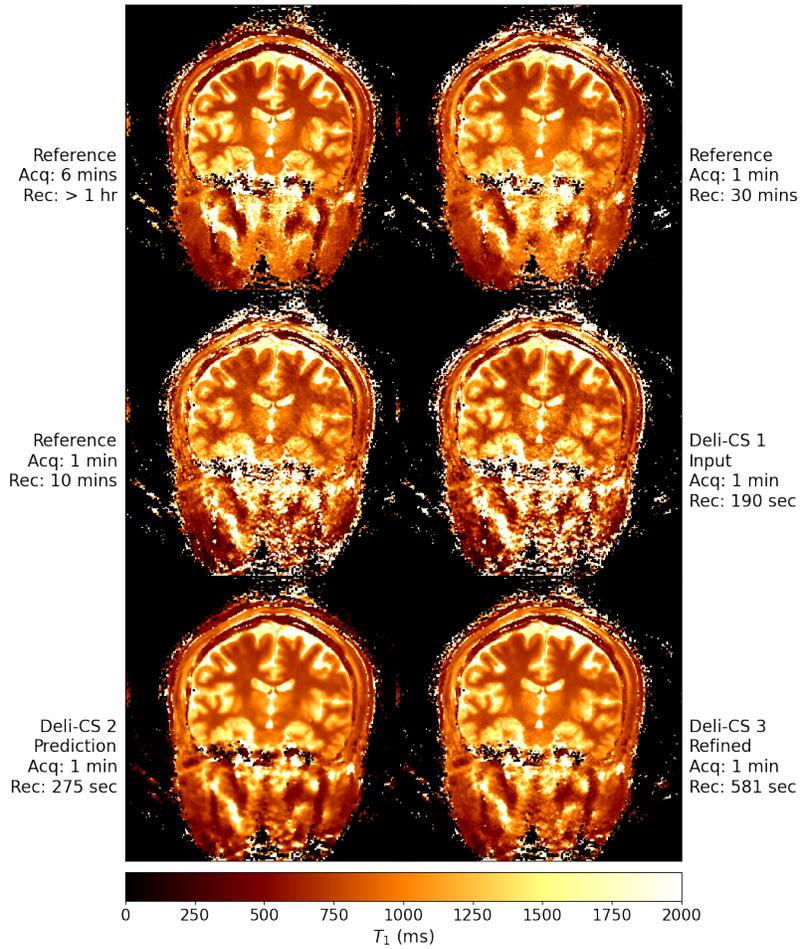


Figure 4-13: This figure compares the  $T_1$  parameter fittings of TGAS-SPI-MRF reconstructions depicted in Figure (4-8). “Acq” denotes the acquisition time, and “Rec” denotes the reconstruction (or inference, for Deli-CS Prediction) time. Note that the Deli-CS timings reported include the previous steps.

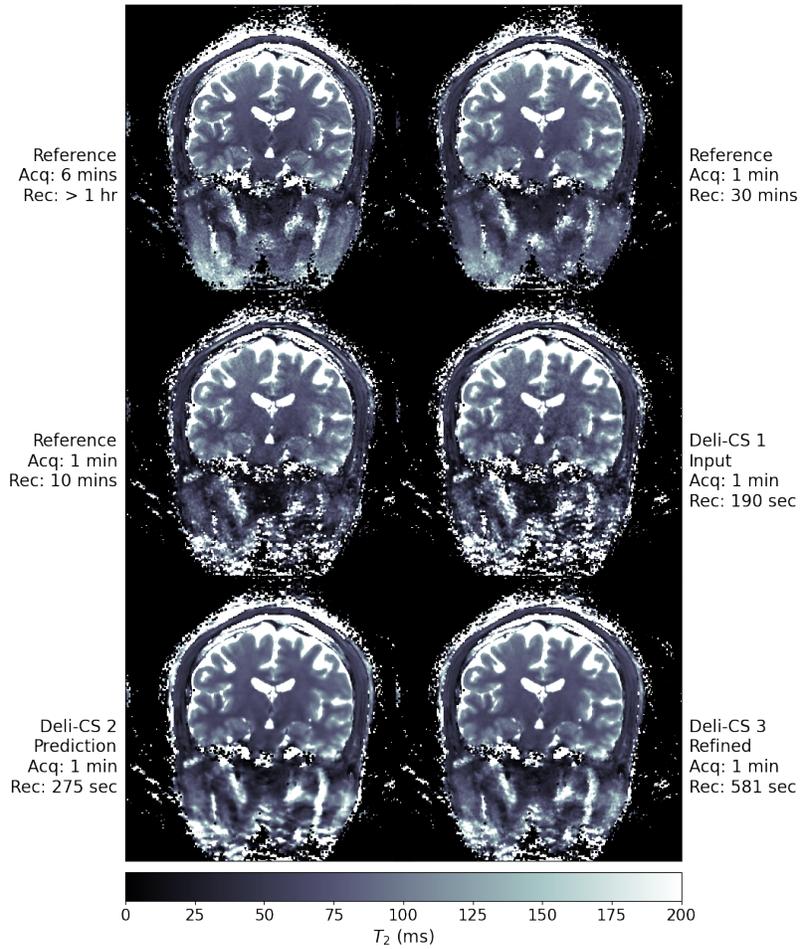


Figure 4-14: This figure compares the  $T_2$  parameter fittings of TGAS-SPI-MRF reconstructions depicted in Figure (4-8). “Acq” denotes the acquisition time, and “Rec” denotes the reconstruction (or inference, for Deli-CS Prediction) time. Note that the Deli-CS timings reported include the previous steps.

average out blocking artifacts from LLR. Consequently, utilizing the polynomial preconditioner reduced the number of shifts performed given the same number of  $A^*DA$  evaluations, which was not desirable. That being said, when performing an iterative reconstruction using the ADMM sub-problem approach in Section (3.4.4), the proximal term in (3.36a) results in a slight Tikhonov regularization (from the ADMM step size) when solving the respective sub-problems which helps alleviate the artifacts induced by the reduced number of shifts over iterations. Assuming multiple 20 GB GPUs, the polynomial preconditioner was shown to enable an approximately 12 minute reconstruction in Section (3.4.4).

While this work uses a simple deep learning model (ResNet[44]) to jump-start the compressed sensing reconstruction, given the flexibility of the Deli-CS framework, various deep learning architecture can be easily integrated to try and improve the quality of the initialization, which is a promising avenue for further optimization. Future work is to leverage a more advanced calibration scan, such as PhysiCal (further discussed in Chapter 6) which aims to acquire  $B_0, B_1^+$  and coil sensitivity information in close to 10 seconds. The  $B_1^+$  map will enable robust parametric mapping, and the  $B_0$  information can be incorporated into the  $A$  matrix in (4.2) to alleviate blurring issues related to spiral imaging that is still present even in the highly accelerated spiral trajectory, particularly in regions where the  $B_0$  is large [15]. This is a good fit for the Deli-CS framework, as the network can learn to predict a  $B_0$  corrected image using the non- $B_0$  corrected CG input, which will result in fewer refinement iterations (that will use the full forward model  $A$  with incorporated  $B_0$ ). This is beneficial as  $A$  augmented with  $B_0$  is computationally challenging as discussed in [15], making the traditional LLR reconstruction even harder to perform. Finally, it would be ideal to port Deli-CS to the more efficient C-language using the BART toolbox [97, 10], which is expected to provide at least another 2 – 3x in speed improvement, moving the application towards near real-time reconstruction.



## 5: Better Encoding of Spatio-Temporal MRI using Wave-Shuffling

This chapter switches gears into MRI encoding optimization, where the acquisition process is modified to yield a measurement matrix  $A$  that results in a better condition regularized linear inverse problem. This chapter has section (2.1) and section (2.4) as “prerequisites”.

### 5.1: Introduction

A number of widely used MR sequences such as Fast Spin Echo (FSE) [67] and Magnetization-Prepared Rapid Gradient-Echo (MPRAGE) [68] reduce scan time by performing k-space encoding along a multi-echo train, which is the time dimension in Section (2.4.4).

For high resolution acquisitions, signal variations along the long echo trains (from (2.11)) can cause significant image blurring. To solve this, spatio-temporal model-based techniques (as discussed in Section (2.4)) have emerged to improve image sharpness while also recovering additional multiple image contrasts across the echo train.

An inherent challenge in high resolution spatio-temporal imaging is acquiring enough k-space samples for each echo to resolve both spatial and temporal dimensions in a clinically feasible time frame. In other words, acquiring enough data such that the  $A$  operator in (2.24) is fully-determined is not practical.

To overcome this, subspace methods have been introduced to greatly reduce the dimensionality of the temporal domain during reconstruction [55]. Despite the smaller dimensionality, the problem is still often ill-posed. Consequently, subspace methods have been used with random sampling and sparsity constrained optimization for better time-resolved imaging and accelerated  $T_1$  and  $T_2$  parameter mapping [55, 61, 94, 15, 102, 45, 111, 100, 5, 99].

$T_2$ -shuffling is one such subspace method that has been applied to clinical volumetric FSE (3D-FSE) imaging [94, 93]. By leveraging random under-sampling in the phase-encode ( $y$ ) and partition ( $z$ ) directions, and applying subspace reconstruction with the LLR low-rank constraint, the method time-resolves an image-series that captures the signal evolution across the echo train and avoids image blurring artifacts. By combating image blurring,  $T_2$ -Shuffling has enabled clinical use of 3D-FSE for pediatric knee MRI by producing diagnostic images in 6 – 7 minutes with similar quality to conventional 2D-FSE acquisition that requires 13 minutes of scan time [2]. The Shuffling approach from  $T_2$ -Shuffling has also been applied to the MPRAGE sequence to extract a time-series of  $T_1$ -weighted images [14].

In addition to the use of subspace modeling, new sequence-based approaches have also been proposed to enable fast time-resolved imaging. MPnRAGE [48] is a multi-echo MPRAGE-like acquisition that leverages radial sampling and a sliding window reconstruction to recover the image time-series across the echo train. Echo Planar Time Resolved Imaging (EPTI) [102] improves on conventional multi-shot Echo Planar Imaging by optimizing k-space and time sampling to exploit temporal correlations during reconstruction to recover distortion free, multi-contrast time series of EPI images. Further more, subspace modeling has been incorporated into EPTI to achieve higher rates of acceleration [32].

In the pursuit of rapid, high resolution spatio-temporal imaging, advances in parallel imaging that reduce the number of k-space samples needed to spatially resolve the underlying image in non time-resolved acquisitions can be leveraged to improve the encoding of accelerated time-resolved imaging. Wave-CAIPI [7] is a parallel imaging technique that utilizes additional sinusoidal gradients during the

readout to spread aliasing in the readout direction. This is shown to improve the acceleration capabilities of parallel imaging through better utility of 3D coil-sensitivity information, resulting in a better posed inverse problem with excellent g-factor performance. Wave-CAIPI has been successfully applied to volumetric 3D-FSE to achieve nine-fold acceleration at  $3T$  [81], and to MPRAGE to enable nine-fold and twelve-fold acceleration at  $3T$  and  $7T$  respectively [81, 82].

However, conventional Wave-CAIPI acquisitions cannot fully utilize the better conditioning provided by wave-encoding. Simulation studies have shown that g-factor performance is largely determined by the maximum gradient amplitude ( $G_{\max}$ ) of the sinusoidal wave-gradient and is independent of the number of sinusoidal cycles [81]. To achieve the maximum possible  $G_{\max}$ , lower wave cycle numbers are required due to gradient hardware slew rate limitations. When utilizing such encoding, conventional Wave-CAIPI acquisitions, where the data are acquired across a long echo train, suffer from undesirable image ringing artifacts due to the interaction between signal modulation along the echo train from  $T_1/T_2$  signal relaxation and the voxel spreading effect from wave-encoding [81]. This ‘‘Signal-Mixing’’ artifact becomes significant when a high  $G_{\max}$  and a low cycle number wave-encoding is used, limiting the amount of encoding performance that can be practically achieved. However, this can be overcome by time-resolving the echo-train during reconstruction. In particular, while Wave-CAIPI was developed as a natural extension to controlled aliasing methods like Bunch Phase Encoding [65] and 2D-CAIPIRINHA [12, 13], the sinusoidal gradients have also been successfully applied to data acquired with random k-space sampling to achieve high rates of acceleration with good reconstruction [6, 20]. This suggests that wave-encoding can be integrated into spatio-temporal model-based techniques to be robust to Signal-Mixing artifacts while improving the acceleration performance of time-resolved applications.

When utilizing wave-encoding, hardware imperfections can cause errors in the sinusoidal gradient resulting in detrimental image ringing artifacts if not accounted for in the reconstruction. Auto-calibrated PSF (or ‘‘AutoPSF’’) methods [17, 20] have been proposed to estimate these gradient imperfections directly from the undersampled data, but these approaches cannot be directly applied to acquisitions with random under-sampling without a fully sampled calibration region due to high computational cost. Consequently, to avoid the need to acquire additional calibration data, a computationally efficient wave-calibration method for the mentioned conditions is required.

In order to fully utilize the better conditioning provided by high  $G_{\max}$  wave-encoding, ‘‘Wave-Shuffling’’ is proposed where wave-encoding is incorporated into the Shuffling model with the temporal subspace and random undersampling. By temporally-resolving the underlying signal during the reconstruction, ‘‘Wave-Shuffling’’ is shown to be robust to the ‘‘Signal Mixing’’ artifacts, enabling higher  $G_{\max}$  compared to conventional Wave-CAIPI. The  $G_{\max}$  of wave-encoding is then optimized to achieve significantly higher acceleration capability compared to standard Shuffling while retaining the multi-contrast and clinically desirable blur-free reconstruction provided by Shuffling [2]. To enable Wave-Shuffling application that is robust to hardware imperfections, a computationally efficient modification of the AutoPSF approach [17] is proposed.

## 5.2: Theory

In Wave-Shuffling, the spatio-temporal model (2.26) is augmented with Wave-encoding (2.20), and is solved akin to (2.27).

$$\operatorname{argmin}_m \frac{1}{2} \|MF_{yz}WF_xRS\Phi\alpha - y\|_2^2 + \lambda g(\alpha) \quad (5.1)$$

As in  $T_2$ -Shuffling, the prior utilized for reconstruction ( $g$  in both (2.27) and (5.1)) is the LLR constraint [94]. The conditioning benefits of the Wave-encoding allows for better utilization of the coil sensitivity information ( $S$ ) for higher under-sampling compared to standard Shuffling.

### 5.2.1 Wave-Encoding Parameter Optimization

Better noise performance is achieved using sinusoidal wave-gradients with higher  $G_{\max}$  which require lower wave cycle numbers due to gradient slew rate limitations. When utilizing such encoding, conventional Wave-CAIPI acquisitions suffer from ringing artifacts, denoted ‘‘Signal-Mixing’’, due to the interaction between the  $T_1/T_2$  signal relaxation over the echo-train and the voxel-spreading effect from wave-

encoding [81]. Wave-Shuffling mitigates this artifact as data are time-resolved across the temporal dimension. Consequently, more efficient wave-encoding is achieved by optimizing the wave-gradient maximum amplitude.

When using LLR, the expected performance of (2.27) for a given Shuffling acquisition is qualitatively analyzed by studying its Point Spread Function (PSF) [94]:

$$\text{PSF}_n = (\Phi^* F_{xyz}^* M F_{xyz} \Phi) \delta_n \quad (5.2)$$

The delta image ( $\delta_n$ ) is zero at all entries except at the center image of the  $n^{\text{th}}$  coefficient, where it is one. The ideal response would be the same delta image for the  $n^{\text{th}}$  coefficient with zero cross talk across the other coefficient images. Due to acceleration and random sampling, the resulting PSF will have noise-like incoherent side-lobes in the  $n^{\text{th}}$  coefficient image with signal aliasing to the other coefficients.

The **peak-to-side-lobe ratio** of  $\text{PSF}_n$  is the largest absolute voxel value over all coefficients of  $\text{PSF}_n$  not including where  $\delta_n$  is non-zero. The lower the peak-to-side-lobe of the PSF over all values of  $n$ , the more likely a LLR reconstruction like (2.27) will “de-noise” the incoherent aliasing artifacts.

The PSF for Wave-Shuffling is calculated as follows.

$$\text{WPSF}_n = (\Phi^* R^* F_x^* W^* F_{yz}^* M F_{yz} W F_x R \Phi) \delta_n \quad (5.3)$$

This is calculated as a function of wave-encoding parameters, and is used as a metric for parameter optimization.

## 5.2.2 Wave-PSF Calibration

Hardware imperfections cause errors in the sinusoidal gradients of wave-encoding, resulting in image ringing artifacts if not accounted for in the reconstruction. Auto-calibrated PSF (or “AutoPSF”) methods [17, 20] estimate these gradient imperfections directly from the under-sampled wave data, but cannot be directly applied to Wave-Shuffling as data are acquired with random undersampling without a fully sampled calibration region. To enable Wave-Shuffling that is robust to hardware imperfections, an efficient AutoPSF method is developed by modifying the approach proposed in [17].

The Wave-PSF ( $W$ ), used in (2.20), is constructed in the  $(k_x, y, z)$  domain [7] as follows:

$$\begin{aligned} W(k_x(t), y, z) &= \exp\{1i \times 2\pi \times [P_y(k_x(t)) \cdot y + P_z(k_x(t)) \cdot z]\} \\ P_y(t) &= \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau \\ P_z(t) &= \frac{\gamma}{2\pi} \int_0^t G_z(\tau) d\tau \end{aligned} \quad (5.4)$$

$(G_y, G_z)$  are the sinusoidal gradients depicted in figure (2-2)(A),  $\gamma$  is the Larmor frequency and  $k_x(t)$  denotes the  $(k_x)$  point sampled at readout-time ( $t$ ). [17] proposed a joint optimization that solves for the underlying image and the imperfections of the sinusoidal gradients during reconstruction where a small set of Fourier coefficients accurately model the sinusoidal wave gradients along with the related gradient hardware errors. Let  $(c_y, c_z)$  denote the small set of Fourier coefficients of the  $(G_y, G_z)$  gradients respectively. Let  $F$  denote the Fourier transform. Then:

$$\begin{aligned} G_y(t) &= [F(c_y)](t) \\ G_z(t) &= [F(c_z)](t) \end{aligned} \quad (5.5)$$

The derived  $(G_y, G_z)$  gradients are used to generate the Wave-PSF ( $W$ ) as in (5.4). Let  $W(c)$  denote the Wave-PSF derived from  $c = (c_y, c_z)$ . AutoPSF alternates between minimizing the image ( $m$ ) and the coefficients ( $c$ ) to reduce the data consistency error:

$$\begin{aligned} \text{Data Consistency: } m &= \operatorname{argmin}_m \|v - M F_{yz} W(c) F_x R S m\|_2 \\ \text{Fourier Update: } c &= \operatorname{argmin}_c \|v - M F_{yz} W(c) F_x R S m\|_2 \end{aligned} \quad (5.6)$$

The first image ( $m$ ) is reconstructed with Fourier estimates ( $c$ ) derived from assuming no  $(G_y, G_z)$  hardware errors.

In the Data Consistency of (5.6), AutoPSF leverages the structured aliasing of Wave-CAIPI for fast computation by selecting a small set of representative voxels from which an estimate of the data-consistency error is calculated. This does not extend to Wave-Shuffling as random sampling causes incoherent aliasing artifacts making it not possible to select a small subset of representative voxels, thus dramatically increasing computational cost. To overcome this, the following PSF-calibration is proposed.

Let  $c_0$  be the Fourier coefficients of the ideal wave sinusoidal gradients (assuming no errors). Let  $c_*$  be the Fourier coefficients of the corrected wave-gradients obtained from the result of (5.6). Define  $c_e$  as:

$$c_e = (c_* - c_0). \quad (5.7)$$

Let  $W_*$ ,  $W_0$  and  $W_e$  denote the Wave-PSFs derived from  $c_*$ ,  $c_0$  and  $c_e$  respectively. Utilizing the linearity of the Fourier transform, (5.7) and (5.4) imply:

$$\begin{aligned} W_*(k_x(t), y, z) &= \exp\{1i \times 2\pi \times ([F((c_*)_y)](k_x(t)) \cdot y + [F((c_*)_z)](k_x(t)) \cdot z)\} \\ &= \exp\{1i \times 2\pi \times ([F((c_e + c_0)_y)](k_x(t)) \cdot y + [F((c_e + c_0)_z)](k_x(t)) \cdot z)\} \\ &= \exp\{1i \times 2\pi \times ([F((c_0)_y)](k_x(t)) \cdot y + [F((c_0)_z)](k_x(t)) \cdot z)\} \times \\ &\quad \exp\{1i \times 2\pi \times ([F((c_e)_y)](k_x(t)) \cdot y + [F((c_e)_z)](k_x(t)) \cdot z)\} \\ &= W_0 W_e \end{aligned} \quad (5.8)$$

This suggests that the artifacts caused by an incorrect Wave-PSF can be modeled by a convolution with  $W_e$ . It may be possible to estimate a deconvolution to recover  $W_*$  using (5.8), which re-casts wave-calibration into a deconvolution problem. Let  $(\alpha_0)$  be the initial Wave-Shuffling reconstruction performed using  $(W_0)$ . This is expected to have ringing artifacts due to system hardware imperfections. It is assumed that these artifacts are well approximated by a convolution with the Wave-PSF  $(W_e)$ , and finding the inverse PSF that mitigates these ringing artifacts allows for the recovery of  $(W_*)$ . In particular, since  $(W_e)$  is a phase only PSF, the inverse PSF is simply the element wise complex conjugate PSF  $(W_e^*)$ . Consequently, finding a PSF that mitigates the ringing artifacts of the initial reconstruction  $(\alpha_0)$  is expected to be a good approximation of  $(W_e^*)$  and consequently  $(W_e)$ .

Note that hardware error related artifacts are exaggerated by applying an edge detector along the readout direction. Let this edge detector be  $T$ . Consider the following optimization problem:

$$c_e = \underset{c}{\operatorname{argmin}} \|T(F_x^{-1} W_e^*(c) F_x \alpha_0)\|_1 \quad (5.9)$$

This optimization attempts to correct for Wave-PSF errors by performing deconvolution as a post-processing step (by applying  $F_x^{-1} W_e^*(c) F_x$ ). After deconvolving  $(\alpha_0)$  with  $W_e^*(c)$ , the edge detector is applied to amplify the remaining artifacts. Since there are only a limited number of coefficients  $(c)$ , (5.9) is efficiently solved with the Nelder-Mead simplex algorithm, and is expected to recover the desired difference coefficients  $(c_e)$ . Other values of  $(c)$  in (5.9) are induce additional undesired voxel-spreading. The corrected Wave-PSF  $(W_*)$  can then be recovered as in (5.8) to be used in a Wave-Shuffling reconstruction.

In this work, the following edge detector is used. Let  $D_x$  denote the finite difference along the readout direction. Let  $\mu$  be a positive scalar.

$$T(m) = \operatorname{sigmoid}(\mu \times |D_x(m)|) - \frac{1}{2} \quad (5.10)$$

$\mu = 100$  is empirically found to sufficiently amplify wave-related ringing artifacts while not being over-sensitive to noise.

For additional computational efficiency, the separability of wave-encoding is utilized. Since  $(G_y, G_z)$  induce voxel spreading independently,  $W$  can be split into separate  $(k_x, y)$  and  $(k_x, z)$  PSFs.

$$W(k_x, y, z) = W_y(k_x, y) W_z(k_x, z) \quad (5.11)$$

To calibrate for  $W_y$  without being affected by  $W_z$ , (5.9) is applied to a single slice of the three-dimensional volume at  $z = 0$  spatial position. The similar process is repeated for  $W_z$  by extracting the slice at  $y = 0$  spatial position.

For succinctness, this method of Wave-Calibration is called ‘‘ShflPSF’’.

## 5.3: Methods

First, experiments to select that rank of the subspace and the wave parameter are performed. This is then followed by in-vivo experiments to verify the efficacy of the method.

### 5.3.1 Subspace Design

The methodology used to generate the low-rank temporal basis is first described.

For all MPRAGE experiments, the Bloch equation was used to simulate the signal evolution across the echo train during the inversion recovery using a flip angle of  $9^\circ$  and  $T_1$  values uniformly sampled across the range [50,5000] milliseconds at 10 millisecond intervals. These are the expected  $T_1$  values of tissues in the brain, and are similar to those used in prior work [108]. To be robust to  $B_1^+$  variation, the simulations were repeated for flip angles across the range [6.3,11.7] degrees at  $0.9^\circ$  intervals. Unless otherwise specified, an echo spacing of 8.1 ms was assumed.

The Singular Value Decomposition (SVD) of the simulated signal ensemble was used to derive the temporal subspace. The ensemble was projected onto the subspace spanned by the first  $k$  singular vectors corresponding to the  $k^{th}$  largest singular values, and the normalized root mean squared error (NRMSE) for each signal was calculated. The maximum NRMSE over all signals was used as a metric to determine the basis rank.

For the 3D-FSE experiment, note that a variable flip angle train was utilized to extend the signal duration over the echo-train [67], allowing signal recovery at echo-times much longer than conventional spin-echo imaging. The signal evolutions were simulated using Extended Phase Graphs [104]. The flip angle train used during acquisition was saved and the simulations were generated assuming  $T_1$  values of [200,400,800,1400,2000,4000] in milliseconds and  $T_2$  values in the range [20,2000] milliseconds at 2 milliseconds intervals. These are the expected  $T_1$  and  $T_2$  values of tissues in the brain and are similar to those used in prior work [108]. FSE has a weak dependence on  $T_1$  [94], hence a smaller range of  $T_1$  values were used. The first 45 echoes (of 256) were discarded as they significantly ill-conditioned the reconstruction. The SVD of the simulated signal ensemble was used to derive the subspace. The signal ensemble was projected onto the subspace spanned by the first  $k$  singular vectors corresponding to the  $k^{th}$  largest singular values and the NRMSE for each signal in the ensemble was calculated. The maximum NRMSE over all signals was used as a metric to determine the number of basis vectors to use.

### 5.3.2 PSF Analysis and Wave-Encoding Parameter Optimization

To examine the benefit of incorporating wave-encoding into the Shuffling, the PSF of MPRAGE Shuffling and MPRAGE Wave-Shuffling was calculated for a k-space sampling mask where 25% of all  $(k_y, k_z)$  encodes were sampled uniformly at random (each encode was sampled exactly once). The matrix size was  $256 \times 256 \times 256$  with a turbo factor (TF) of 256. The PSFs calculations were implemented using MATLAB (MathWorks, Natick, MA). To aid visualization, only two basis vectors (corresponding to the two largest singular values) are used.

Delta images  $\delta_1$  and  $\delta_2$  were passed through the Shuffling and Wave-Shuffling forward models ((5.2) and (5.3)) to obtain “PSF<sub>1</sub>” and “PSF<sub>2</sub>” for Shuffling, and “WPSF<sub>1</sub>” and “WPSF<sub>2</sub>” for Wave-Shuffling respectively.

The peak-to-side-lobe ratios of the PSFs were calculated. The result of the following (5.12) as a function of wave cycle number and maximum wave amplitude are visualized.

$$\max_n [\text{peak-to-side-lobe}(\text{WPSF}_n)] \quad (5.12)$$

This value was used to optimize the parameters for wave-encoding, as discussed in Section (5.2.1) For a given cycle number, the maximum amplitude was determined by either gradient hardware slew rate limitations or the peripheral nerve stimulation limit.

### 5.3.3 In-Vivo Experiments

The 3D-FSE and MPRAGE sequences were modified to enable sinusoidal wave-gradients and random sub-sampling of k-space. Three-dimensional Standard Wave MPRAGE, MPRAGE Shuffling, MPRAGE Wave-Shuffling and 3D-FSE Wave-Shuffling data from a single 24 year old healthy male were acquired on a 3T Siemens Prisma scanner with IRB approval and informed consent. All data were acquired using a matrix size of  $256 \times 256 \times 256$ , 1 mm-isotropic resolution, and 32-channel head coil. The turbo factor for both MPRAGE and 3D-FSE was set to 256. 16-channel SVD coil compression was applied to all data sets. Coil sensitivity maps were estimated using ESPIRiT [96] from a 2-second low-resolution GRE calibration scan. All Shuffling and Wave-Shuffling data were prospectively under-sampled using a variable density Poisson disc sampling mask generated using BART [97]. Each phase encode was sampled at most once. The Standard Wave MPRAGE data were acquired using a regular  $2 \times$  undersampling pattern along  $k_y$ .

This work lists the total acquisition times ( $T_{acq}$ ) instead of acceleration factors. An acquisition time of 648 seconds corresponds to a fully sampled non time-resolved MPRAGE acquisition with the same sequence parameters.

For all Standard Wave acquisitions, the associated Wave-PSFs were calibrated using [17]. For the Wave-Shuffling acquisitions, the respective Wave-PSFs were estimated using ShflPSE. A 50-iteration reconstruction using a single basis was used as the initial reconstruction ( $\alpha_0$ ) in (5.9).

The Wave-Shuffling reconstructions were performed at a synthetic resolution of  $\approx 0.71$  mm in ( $y, z$ ) directions by zero-padding the ( $k_y, k_z$ ) dimensions. The sparsifying transform used was LLR. The reconstruction algorithm used FISTA [4]. For all cases, the acquired k-space was normalized to have unitary  $l_2$  norm. Regularization values ( $\lambda$ ) from the below set were searched through for the best qualitative reconstruction, yielding a regularization value of  $\lambda = 0.002$ .

$$\lambda \in \{0.0001, 0.0002, \dots, 0.0009, 0.001, 0.002, \dots, 0.009, 0.01, 0.02\} \quad (5.13)$$

This value was then kept constant over all cases.

The Wave-Shuffling reconstruction was implemented with BART. Additionally, BART was compiled using the Intel(R) Math Kernel Library. The reconstructions were performed on an Intel(R) Xeon(R) Gold 6248R CPU.

#### MPRAGE Experiments

As per [81], acquisitions were performed using an inversion time (TI) of 1100 ms, a repetition time (TR) of 2500 ms and a bandwidth (BW) of 200 Hz/Pixel. The wave-encoding used a sine wave on one gradient axis and a cosine wave on the other gradient axis (see Figure (5-1)(A)). The cosine wave-gradient required a ramp up gradient period to reach the highest gradient amplitude at the start of the acquisition window and a ramp down period after the end of the acquisition to return to zero. For simplicity, the current cosine implementation utilized a quarter-cycle shifted sine wave with an additional half cycle as shown in the  $G_z$  gradient of Figure (5-1)(A).

This increases the minimum viable echo spacing when the size of the wave sinusoid is large, which then necessitates an increase to the minimum TI. The minimum possible number of cycles achievable while maintaining a TI of 1100 milliseconds seconds was five. Consequently, this work used 5 cycles as a lower limit to the number of cycles.

Standard Wave MPRAGE, MPRAGE Shuffling and MPRAGE Wave-Shuffling data were acquired using the parameters listed in Table 5.1. Regular Shuffling data at various prospective accelerations were acquired to determine how much acceleration can be achieved. Wave-Shuffling data at different prospective accelerations were also acquired at 17 cycles with a  $G_{max}$  of 8 mT/m to match the wave parameters used in [81]. To push  $G_{max}$  of the wave-encoding to the maximum value while keeping the same TI, Wave-Shuffling data at various prospective accelerations were acquired at 5 cycles with a  $G_{max}$  of 27 mT/m. To verify the efficacy of the Wave-Shuffling model in mitigating Signal Mixing artifacts, Standard Wave MPRAGE data were acquired at 17 cycles with a  $G_{max}$  of 8 mT/m (to match [81]) and at 5 cycles with  $G_{max}$  of 27 mT/m (to demonstrate the Signal Mixing artifacts).

For each acquisition listed in Table 5.1, the Bloch equation was used to simulate signal evolutions across the echo train during the inversion recovery of MPRAGE using the acquisition parameters de-

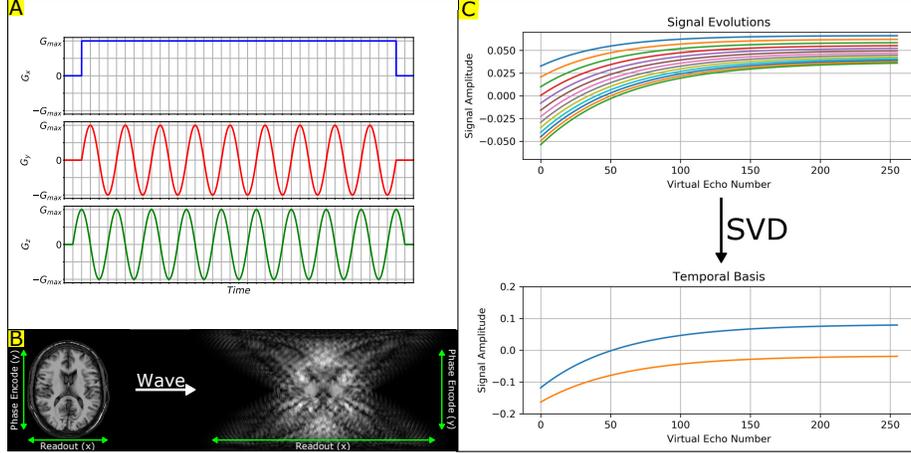


Figure 5-1: (A) In wave-encoding, additional sinusoidal ( $G_y, G_z$ ) gradients are applied during the readout. (B) The sinusoidal gradients induces a voxel spreading effect along the readout direction. The amount of spreading increases linearly as a function of the spatial ( $y, z$ ) locations. (C) This depicts simulated MPRAGE signal evolutions. These realistic signal evolutions are observed to live in a low-rank subspace that can be obtained via the Singular Value Decomposition (SVD).

Acquisition Type	TI (ms)	TR (ms)	ESP (ms)	Cycles	$G_{\max}$ (mT/m)	$T_{acq}$ (s)
Standard Wave	1100	2500	7.9	17	8	324
Standard Wave	1100	2500	8.4	5	27	324
Shuffling	1100	2500	7.8	0	0	648
Shuffling	1100	2500	7.8	0	0	230
Shuffling	1100	2500	7.8	0	0	144
Shuffling	1100	2500	7.8	0	0	81
Wave-Shuffling	1100	2500	7.9	17	8	144
Wave-Shuffling	1100	2500	7.9	17	8	81
Wave-Shuffling	1100	2500	8.4	5	27	144
Wave-Shuffling	1100	2500	8.4	5	27	81

Table 5.1: MPRAGE acquisition parameters. TI: Inversion Time, TR: Repetition Time. ESP: Echo Spacing,  $G_{\max}$ : Maximum amplitude of the sinusoidal gradients,  $T_{acq}$ : Acquisition time.

scribed in the Section (5.3.2) with the appropriate echo-spacing.

All reconstructions were allowed to run for 500 FISTA[4] iterations.

### 3D-FSE Experiment.

For 3D-FSE, two cycle Wave-Shuffling data were acquired at the maximum possible amplitude of 22 mT/m. This corresponds to the lowest number of cycles such that the echo spacing and echo train length of the acquisition is within 20% of prior work [81]. Note that this is a variable flip angle acquisition with stimulated echoes, allowing for the reconstruction of images at later echo-times compared to conventional spin-echo imaging. The echo spacing was 4.32 milliseconds, the TR was 3.2 seconds, the BW was 592 Hz per pixel and the middle echo was acquired at an echo time of 557 milliseconds. The data was acquired with an acquisition time of 151 seconds.

The reconstruction was allowed to run for 500 FISTA [4] iterations with LLR.

## 5.4: Results

This section will first present the results of parameter experiments, as they are used by the in-vivo reconstructions.

### 5.4.1 Subspace Design

For both MPRAGE and 3D-FSE, a rank of 4 was the smallest subspace such that the NRMSE of any signal was less than 2.5% (depicted in Figure (5-2)). The x-axis denotes an index into the signal ensemble (please see caption for a detailed description) and the y-axis denotes the NRMSE.

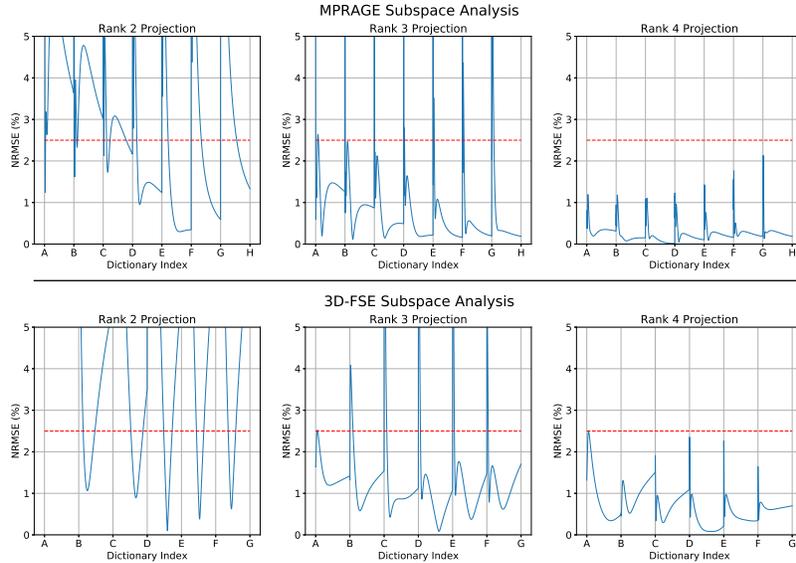


Figure 5-2: The simulated signal ensemble for MPRAGE and 3D-FSE are projected onto SVD-derived subspaces of varying ranks and the NRMSE per signal is calculated. The red dotted line denotes the 2.5% cut-off. For MPRAGE, the range (A-B) corresponds to  $T_1$  values in range [50, 5000] milliseconds with a flip angle of  $6.3^\circ$ , (B-C) corresponds to the same  $T_1$  values with a flip angle of  $7.2^\circ$ , (C-D) corresponds to the same  $T_1$  values with a flip angle of  $8.1^\circ$ , (D-E) corresponds to the same  $T_1$  values with a flip angle of  $9^\circ$ , (E-F) corresponds to the same  $T_1$  values with a flip angle of  $9.9^\circ$ , (F-G) corresponds to the same  $T_1$  values with a flip angle of  $10.8^\circ$ , and (G-H) corresponds to the same  $T_1$  values with a flip angle of  $11.7^\circ$ . For 3D-FSE, the range (A-B) corresponds to  $T_2$  values in range [20, 2000] milliseconds with a  $T_1$  of 200 ms, (B-C) corresponds to the same  $T_2$  values with a  $T_1$  value of 400 ms, (C-D) corresponds to the same  $T_2$  values with a  $T_1$  value of 800 ms, (D-E) corresponds to the same  $T_2$  values with a  $T_1$  value of 1400 ms, (E-F) corresponds to the same  $T_2$  values with a  $T_1$  value of 2000 ms, and (F-G) corresponds to the same  $T_2$  values with a  $T_1$  value of 4000 ms.

### 5.4.2 PSF and Parameter Optimization

PSF results are depicted in Figure (5-3).

The maximum of the peak side-lobes as a function of the maximum sinusoid amplitude ( $G_{\max}$ ) and number of cycles is depicted. For a particular  $G_{\max}$ , a lower peak side-lobe is observed when using lower cycles. For a particular cycle, a lower peak side-lobe is observed when using higher  $G_{\max}$ . Therefore, in the sequel, Wave-Shuffling acquisitions with a low number of cycles and large  $G_{\max}$  are targeted.

The PSFs of MPRAGE Shuffling (no wave encoding) and MPRAGE Wave-Shuffling with the highest  $G_{\max}$  of 27 mT/m are plotted. The maximum along the  $(x, y)$ ,  $(x, z)$  and  $(y, z)$  spatial axes are taken to better visualize the incoherence. “PSF<sub>1</sub>” and “<sub>2</sub>” depict the Shuffling outputs of  $\delta_1$  and  $\delta_2$  respectively. “WPSF<sub>1</sub>” and “WPSF<sub>2</sub>” depict the Wave-Shuffling outputs of  $\delta_1$  and  $\delta_2$  respectively. The green dashed

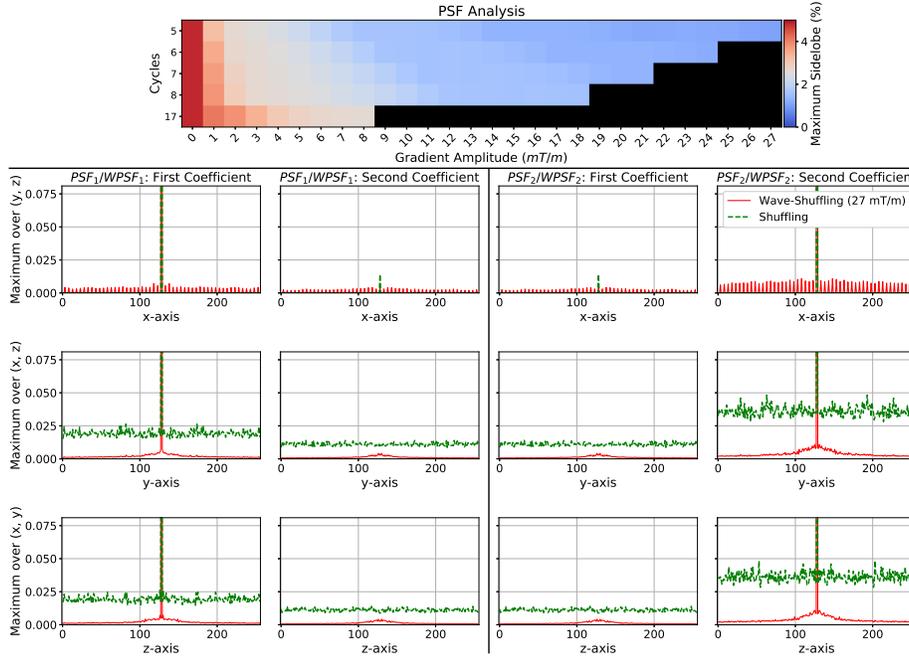


Figure 5-3: (Top) The maximum peak side-lobe as a function of sinusoid maximum gradient amplitude ( $G_{\max}$ ) and number of cycles is depicted. (See Equation (5.12).) A lower peak side-lobe is desirable. The black color denotes unreachable parameters due to gradient hardware slew rate limitations. (Bottom Left and Bottom Right) The PSFs associated with MPRAGE Shuffling (with no wave encoding) and high  $G_{\max}$  MPRAGE Wave-Shuffling are plotted for comparison. The former is plotted with a dotted green line, while the latter is plotted with a solid red line. The legend of the PSFs plots are depicted in the top-right sub-plot. The left two columns denote the first coefficient response while the right two columns denote the second coefficient response. To better visualize the side lobes, the peaks of the respective PSFs are scaled to one and the range  $[0, 0.1]$  is plotted. The maximum intensity projection along the  $(x, y)$ ,  $(x, z)$  and  $(y, z)$  spatial axes are taken to better visualize the incoherence. PSF<sub>1</sub>/WPSF<sub>1</sub> and PSF<sub>2</sub>/WPSF<sub>2</sub> depict the Shuffling/Wave-Shuffling outputs of  $\delta_1$  and  $\delta_2$  respectively.

lines denote Shuffling results and the red solid lines denote Wave-Shuffling results. The voxel-spreading induced by Wave distributes the incoherent artifacts across  $(x, y, z)$  resulting in lower peak side-lobes. This demonstrates how Wave-Shuffling achieves less signal cross-talk between coefficients.

### 5.4.3 In-Vivo Experiments

Figure (5-4) depicts the reconstructions at selected TIs for MPRAGE Shuffling at various levels of acceleration with  $T_{acq} = (648s, 230s, 144s)$ , and MPRAGE Wave-Shuffling at high acceleration ( $T_{acq} = 81s$ ). For ease of visualization, each image is individually normalized. This figure illustrates how Wave-Shuffling provides comparable reconstruction at  $\approx 8\times$  higher acceleration compared to Shuffling. MPRAGE Shuffling is stable at  $T_{acq} = 648s$  but suffers from reconstruction artifacts at higher accelerations that are particularly noticeable at the early time points. In contrast, MPRAGE Wave-Shuffling at  $T_{acq} = 81s$  is able to stably recover the underlying time series with reconstruction quality comparable to MPRAGE Shuffling at  $T_{acq} = 648s$  up-to noise arising from high levels of acceleration.

Figure (5-5) demonstrates the reconstruction quality of Wave-Shuffling as a function of wave-encoding parameters, where reconstructions from a low-SNR TI with low signal level are shown to highlight the differences in reconstruction performance. Regular Shuffling has severe artifacts at the displayed TI at both accelerations  $T_{acq} = 144s$  and  $T_{acq} = 81s$ . The 17-cycle,  $G_{\max} = 8$  mT/m case is seen to improve reconstruction at  $T_{acq} = 144s$  and  $T_{acq} = 81s$ . The 5-cycle,  $G_{\max} = 27$  mT/m case is seen to provide much better

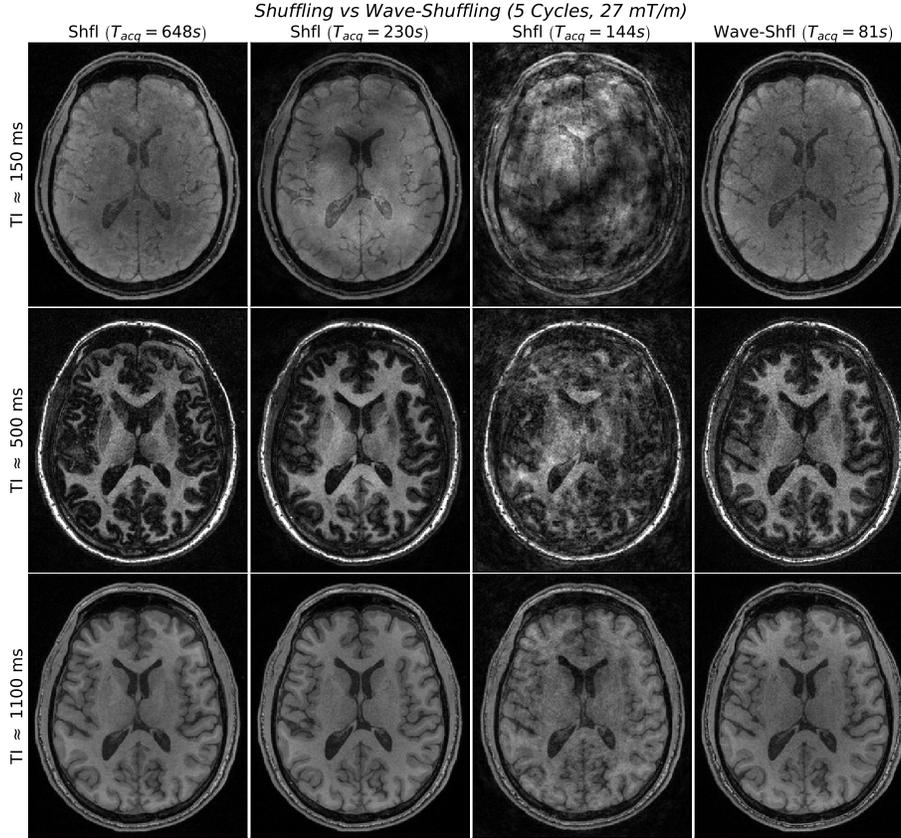


Figure 5-4: This figure depicts MPRAGE Shuffling at various accelerations and MPRAGE Wave-Shuffling at high acceleration. MPRAGE Shuffling is stable at  $T_{acq} = 648s$  but suffers from reconstruction artifacts at lower acquisition times. MPRAGE Wave-Shuffling at a significantly lower acquisition time provides well conditioned reconstruction with image quality comparable to MPRAGE Shuffling at  $T_{acq} = 648s$  up-to noise considerations due to sub-sampling.

conditioning and stably recovers the data at  $T_{acq} = 81s$ .

Figure (5-4) and 5-5 demonstrate how optimized wave-encoding enables Wave-Shuffling to achieve significant increase in acceleration. There are slight contrast differences between the image at the specific TIs due to echo spacing differences between the acquisitions (listed in Table 5.1).

Figure (5-6) demonstrates how Shuffling mitigates Signal Mixing artifacts. Standard Wave MPRAGE with parameters used as per [81] (17 cycles,  $G_{max} = 8$  mT/m) shows good reconstruction with no artifacts. Standard Wave MPRAGE at 5 cycles,  $G_{max} = 27$  mT/m suffers from significant ringing artifacts from the signal recovery mixing across the echo-train. Wave-Shuffling at the same wave parameters (5 cycles,  $G_{max} = 27$  mT/m) at the comparable TI of 1100 milliseconds shows no ringing artifacts even at high acceleration ( $T_{acq} = 81s$ ). This confirms that modelling the temporal evolution over the echo-train avoids the Signal Mixing.

Figure (5-7) demonstrates how ShflPSF corrects for gradient hardware errors and mitigates ringing artifacts related to incorrect wave sinusoids by performing a post-processing deconvolution with  $W_e^*$  as described in the Section (5.2.2).

Figure (5-8) represents a showcase of the Wave-Shuffling method when applied to the MPRAGE sequence for 1-mm isotropic resolution at  $T_{acq} = 81s$ . Four out of the 256 reconstructed images are depicted. A highly accelerated MPRAGE Wave-Shuffling acquisition and reconstruction stably recovers a time series of images with multiple contrasts at high quality up to noise considerations. Note that the CSF is hypo-dense compared to relative tissue, particularly at short TE. Due to having a high  $T_1$  value, the CSF

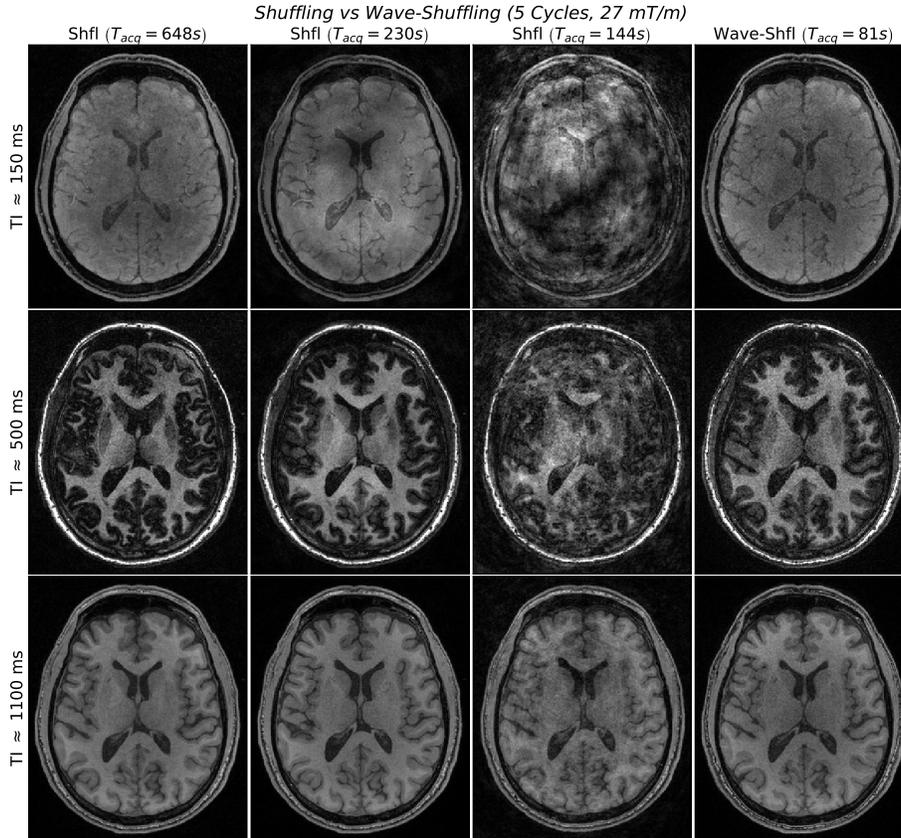


Figure 5-5: The reconstruction quality of Wave-Shuffling varies as a function of wave-encoding parameters. Regular Shuffling is seen to have severe artifacts at the hard-to-reconstruct displayed time point (TI = 430 ms) at both accelerations ( $T_{acq} = 144s$  and  $T_{acq} = 81s$ ). The 5-cycle,  $G_{max}$  of 27 mT/m case demonstrates the superior encoding capability of high  $G_{max}$  wave-encoding with significantly better reconstruction quality at both accelerations.

is unable to fully recover within the prescribed TR and consequently, once the signal reaches steady-state, the CSF signal is suppressed relative to the surrounding tissue.

Figure (5-9) represents a showcase of the Wave-Shuffling method when applied to the 3D-FSE sequence for 1-mm isotropic acquisition at  $T_{acq} = 151s$ . Four out of the 256 reconstructed images are depicted. A highly accelerated 3D-FSE Wave-Shuffling acquisition and reconstruction recovers a time series of images with multiple contrasts. Since 3D-FSE uses a variable flip angle train to stimulate echoes, there is still signal at around 500 ms.

## 5.5: Discussion

In this chapter, Wave-Shuffling was developed for fast time-resolved structural imaging. Wave-encoding parameters were optimized using the PSF analysis to improve acceleration capability. Wave-Shuffling was successfully implemented on MPRAGE and 3D-FSE sequences and demonstrated to provide fast high-quality time-resolved brain imaging.

A direct application of the Wave-Shuffling technique is the further reduction of scan times for knee MRI exams [93]. A promising direction of Wave-Shuffling MPRAGE is to study deep-brain structure like the lateral geniculate nucleus which is more visible at the earlier TIs [28]. This is motivated by the quality of the reconstruction performance demonstrated by Figure (5-8) with just 81 seconds of acquisition. The SNR is expected to significantly improve at scan times comparable to MP2RAGE [63, 69]. Note that

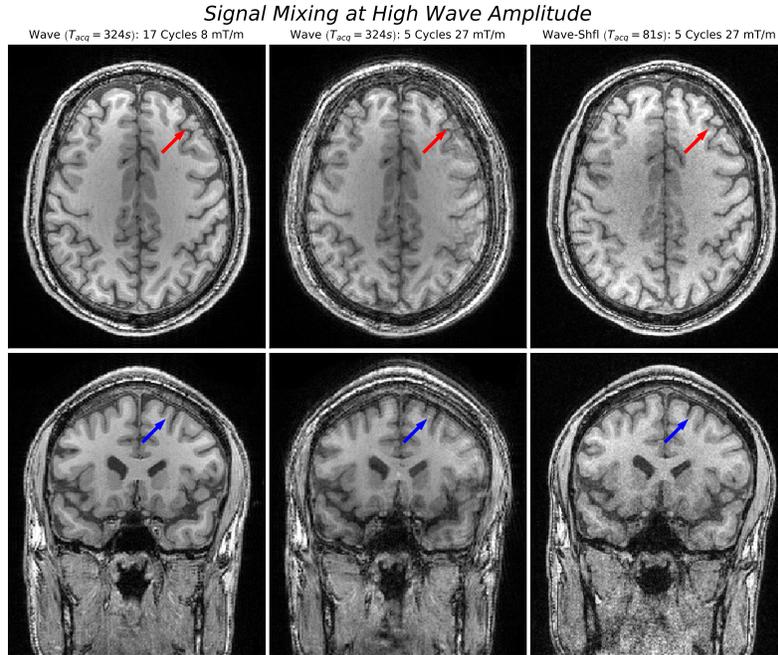


Figure 5-6: The inclusion of the Shuffling model is seen to mitigate Signal Mixing artifacts associated with high gradient amplitude ( $G_{\text{max}}$ ) wave-encoding acquisitions. Standard Wave MPRAGE at 17 cycles,  $G_{\text{max}}$  of 8 mT/m shows good reconstruction with no artifacts. Standard Wave MPRAGE at 5 cycles,  $G_{\text{max}}$  of 27 mT/m suffers from significant ringing artifacts due to signal recovery over the partition encode direction. MPRAGE Wave-Shuffling at the same parameters (5 cycles,  $G_{\text{max}}$  of 27 mT/m) at the comparable TI of 1100 milliseconds shows no ringing artifacts even at high acceleration. For high  $G_{\text{max}}$  Standard Wave MPRAGE (the middle column), the red and blue arrows highlight two different locations where the Signal-Mixing artifacts are observed. For the other two acquisitions, the arrows point to the respective comparable locations to demonstrate that low  $G_{\text{max}}$  Standard Wave MPRAGE and high  $G_{\text{max}}$  MPRAGE Wave-Shuffling are robust to said artifacts.

Wave-Shuffling MPRAGE in this work resolves 256 echo-images versus the 2 echo-images recovered by MP2RAGE.

Figure (5-2) shows that the SVD-derived subspace has difficulty representing low  $T_1$  values at lower ranks for MPRAGE and low  $T_2$  values for 3D-FSE. While higher ranks results in lower projection NRMSE, the added unknowns ill-conditions the reconstructions described by Equations (2.27) and (5.1).

For MPRAGE Wave-Shuffling, the lower bound on the number of wave cycles was set to five to maintain a TI of 1100 ms. A four cycle acquisition corresponds to a TI of 1130 ms and a three cycle acquisition corresponds to a TI of 1170 ms. Lower wave cycle numbers achieves higher  $G_{\text{max}}$  but may result in differential phase modulation within a voxel resulting in intra-voxel de-phasing [7]. At three cycles, the MPRAGE Wave-Shuffling reconstruction suffers from significant artifacts. With these constraints in mind, the PSF results in Figure (5-3) are a good qualitative indicator of Wave-Shuffling reconstruction performance. Similarly, for 3D-FSE, one cycle wave-encoding results in the non-trivial artifacts.

The reconstruction noise level is tied to the amount of acceleration. Wave-Shuffling, being a time-resolving method, is encoding-limited and SNR-limited (due to less noise averaging) at high accelerations, and consequently the reconstructions presented in Figures (5-4) to (5-6) show increased noise levels. Areas to explore in this regard are optimized flip-angle encoding for a more noise-performant acquisition and machine-learning based de-noising.

ShflPSF mitigates the ringing artifacts associated with an incorrect Wave-PSF. It is empirically observed that a single coefficient usually suffices to effectively correct for the ringing artifacts. Since the Fourier coefficient is allowed to be complex, the magnitude of the Fourier coefficient effectively models

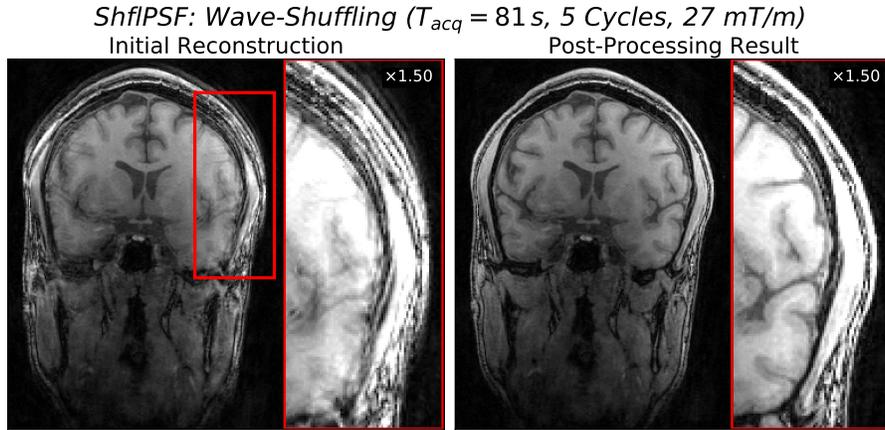


Figure 5-7: This figure depicts ShflPSF applied to the 5 cycles, 27 mT/m maximum gradient amplitude MPRAGE Wave-Shuffling acquisition at  $T_{acq} = 81\text{ s}$ . ShflPSF is able to correct for gradient hardware errors and mitigate ringing artifacts related to incorrect wave calibration by performing deconvolution with  $W_e^*$  as described in the Section (5.2.2).

gradient amplitude errors while the phase of the coefficient encodes gradient delay errors.

Multiple rounds of ShflPSF that alternates between reconstruction and Wave-PSF calibration can be performed. However, this does not noticeably improve reconstruction performance. For Standard Wave acquisitions, ShflPSF performs similarly to [17].

MPRAGE and 3D-FSE are commonly used sequences for  $T_1$ - and  $T_2$ - weighted images respectively that are designed for maximizing contrasts differences between tissues and not for quantitative mapping. The signal evolutions over the echo train are highly correlated for both sequences. To give a specific example, for MPRAGE, assuming a flip angle of  $9^\circ$ , the absolute correlation coefficient between two simulated signals with  $T_1$  values of 800 ms and 1000 ms is  $\approx 0.99$ . Similarly, for 3D-FSE, assuming a  $T_1$  of 800 ms, the absolute correlation coefficient between two simulated signals with  $T_2$  values of 20 ms and 40 ms is  $\approx 0.98$ . Since this work was focused on augmenting existing sequences with known contrasts without modifying sequence parameters, it is not optimized for  $T_1$  and  $T_2$  mapping as there is no clean separability between the highly correlated signal evolutions. Future work would be to modify the flip angle train and other sequence parameters to increase the separability of the signal evolutions of tissues of interest in order to better fit for quantitative maps.

Given the flexibility of the wave-encoding and Shuffling models, Wave-Shuffling can be extended to numerous other sequences with the possibility of complimentary-sampling and joint-reconstruction across multiple Wave-Shuffling acquisitions for higher rates of acceleration and better noise performance [8], which may be further enhanced with Deep Learning methods [80, 41, 1]. To further improve reconstruction, sampling mask optimization that is synergistic with regularization can be utilized [39].

*MPRAGE Wave-Shuffling (5 Cycles 27 mT/m): 81 seconds*

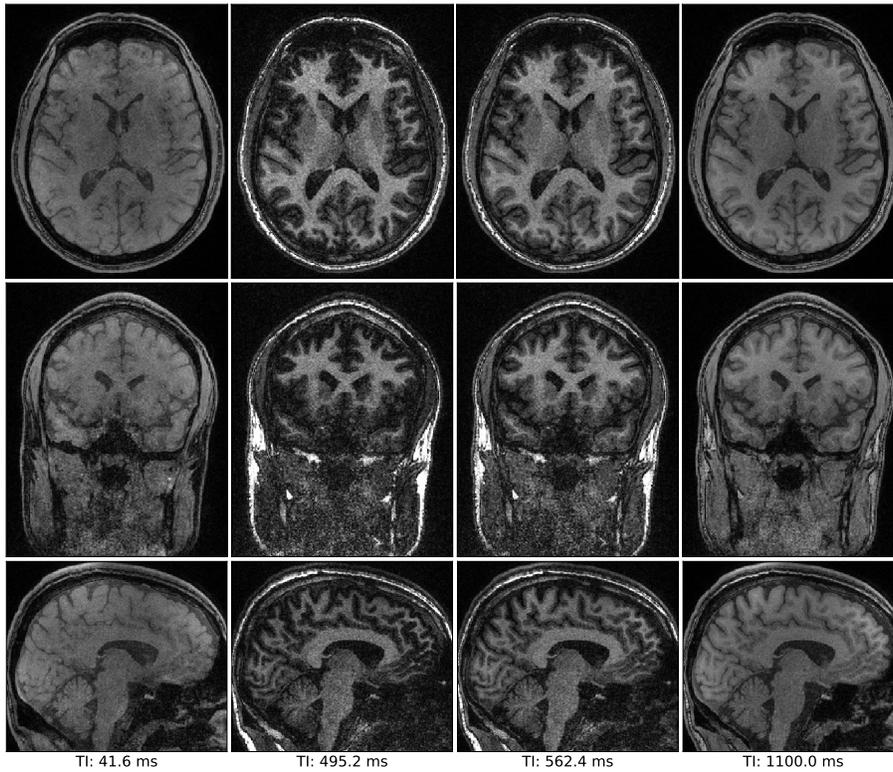


Figure 5-8: Wave-Shuffling achieves 1 mm-isotropic resolution, time-resolved, multi-contrast MPRAGE imaging at high acceleration. Four out of 256 reconstructed images are depicted.

*3D-FSE Wave-Shuffling (2 Cycles 22 mT/m): 151 seconds*

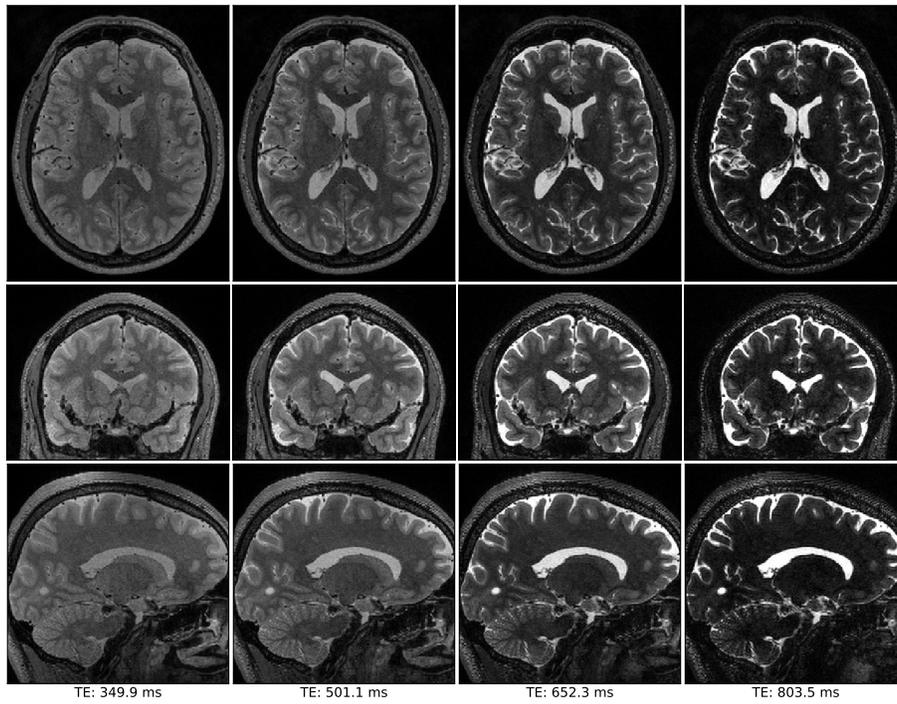


Figure 5-9: Wave-Shuffling achieves 1 mm-isotropic resolution, time-resolved, multi-contrast 3D-FSE imaging at high acceleration. Four out of 256 reconstructed images are depicted.



# 6: Physics Calibration for Spatio-Temporal MRI

This chapter has section (2.3) and section (2.4) as “prerequisites”.

## 6.1: Introduction

Calibration scans for coil sensitivity maps ( $B_1^-$ ),  $B_0$  and  $B_1^+$  inhomogeneities information play an important role in enabling modern acquisition and reconstruction techniques. Great progress has been made in improving the accuracy and speed of these scans. For coil sensitivity estimation, ESPIRiT [96] and JSENSE [109] have been successful in enabling wide-spread parallel imaging, while for accelerated echo planar imaging, FLEET-ACS [83] and Gradient Echo (GRE) with  $B_0$  [110] field map have provided distortion-matched coil sensitivity information for robust reconstruction. For  $B_1^+$  mapping, Bloch-Siebart (BS) methods [87] have gained prominence due to its flexibility and robustness, where recent improvements through k-space undersampling and constrained reconstruction has enabled rapid  $B_1^+$  mapping [53]. Moreover, a robust multi-echo general linear modelling (GLM) framework for BS has also been developed that enables robust recovery of  $B_0$  and  $B_1^+$  maps [22]. Nonetheless, the acquisition of multiple calibration scans for high-resolution coil sensitivity,  $B_0$  and  $B_1^+$  maps can be time consuming, taking 5-10 minutes for whole-brain coverage.

This work proposes a unified, rapid calibration sequence termed Physics Calibration (PhysiCal) to obtain accurate  $B_0$ ,  $B_1^+$  and coil sensitivity maps. PhysiCal utilizes a modified BS multi-echo GRE acquisition, with a carefully designed mix of full and variable density sampling acquisitions across echoes to provide complementary information. This along with synergistic constrained and eigenvalue reconstruction enable significant speedup of this calibration scan. Retrospective undersampling experiments demonstrate robust and accurate recovery of whole-brain  $B_0$ ,  $B_1^+$  and channel coil sensitivity maps in just 11 seconds at resolution at 3T. Furthermore, preliminary verification of PhysiCal is presented through using the rapidly acquired  $B_0$  and  $B_1^+$  maps to process high-resolution diffusion-imaging data acquired with accelerated gSlider-EPI [90, 56].

## 6.2: Methods

The acquisition process is summarized by Figure (6-1). A modified BS, bi-polar, multi-echo 3D-GRE acquisition is used with interleaved, opposite, off-resonant frequencies ( $\pm 4kHz$ ) [54] and with multi-echo readouts prior and subsequent to the BS pulse to achieve estimation robustness.

The k-space sampling of this acquisition is optimized with:

- A small fully-sampled auto-calibrated signal (ACS) in the first echo for ESPIRiT coil sensitivity estimation.
- Independently drawn variable density Poisson-disc sampling across echoes.

This ensures that the residual aliasing artifacts of these image echoes after a regularized reconstruction (1.1) at very high accelerations are incoherent across echoes and can be robustly read-through during the subsequent GLM parameter fitting (analogous to MRF dictionary fitting).

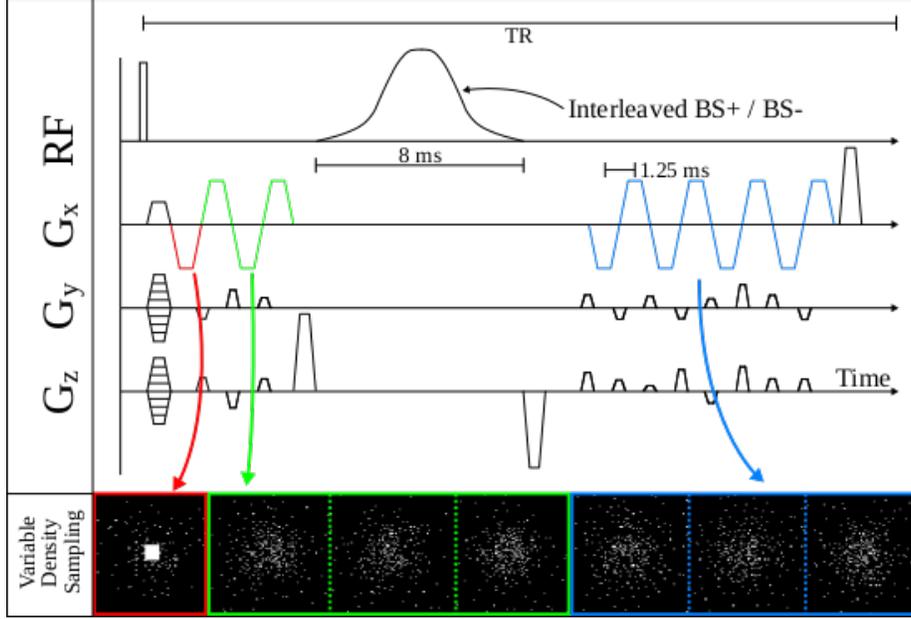


Figure 6-1: A multi-echo GRE sequence is modified to play a strong, Gaussian-shaped, off-resonant RF frequency pulse (denoted BS) after four echoes. The BS pulse is played at alternating opposite off-resonant frequencies ( $\pm 4kHz$ ), denoted as  $BS+$  /  $BS-$ . Each echo is sampled according to an independently drawn variable density Poisson-disc distributed sampling mask, with the first echo consisting of a densely sampled ACS region.

Figure (6-2) depicts the reconstruction process. ESPIRiT is used to calibrate coil sensitivity estimation from ACS of the first echo. This is then used to perform a highly regularized reconstruction using the  $l_1$  norm and sparsity in the Wavelet domain. GLM then robustly recovers artifact-free  $B_0$  maps from reconstructed echo images (which contain temporally-incoherent residual artifacts). However, GLM is seen to not recover an artifact free  $B_1^+$  map.

To overcome this,  $B_1^+$  is refined using an eigenvalue reconstruction approach inspired by ESPIRiT. Phase of the reconstructed echoes after the negative-offset BS pulse is subtracted from the phase of the reconstructed echoes from the positive-offset BS echoes. The resulting time-series is reshaped into “virtual coils” and passed into ESPIRiT, which is expected to recover the smooth underlying  $B_1^+$ .

### 6.3: Experiment

Fully-sampled  $1 \times 2 \times 2 \text{ mm}^3$  resolution PhysiCal data with 16 echoes (4 before BS) are acquired in 19 minutes and 41 seconds with a TR of 36 ms and echo spacing of 1.25 ms. Gold standard  $B_0$  and  $B_1^+$  maps are estimated using GLM and coil sensitivity information is estimated from the first GRE echo using ESPIRiT.

This acquired data are retrospectively undersampled by varying the number of  $(k_y, k_z)$  sampled points per echo and the number of echoes. The first echo contains  $16 \times 16$  ACS for ESPIRiT estimation, with  $B_0$  and  $B_1^+$  estimated using the procedure outlines in Section (6.2). These are then compared to the gold standard. BART[97] was used for sampling-mask generation, ESPIRiT estimation and regularized reconstructions.

To provide preliminary verification of the efficacy and application of the method, retrospectively accelerated parametric maps obtained from PhysiCal are used in EPI-gSlider.  $B_1^+$  inhomogeneity is used to mitigate striping artifacts in an EPI-gSlider acquisition.  $B_0$  is used to perform post-processing distortion correction to un-distort an EPI reconstruction.

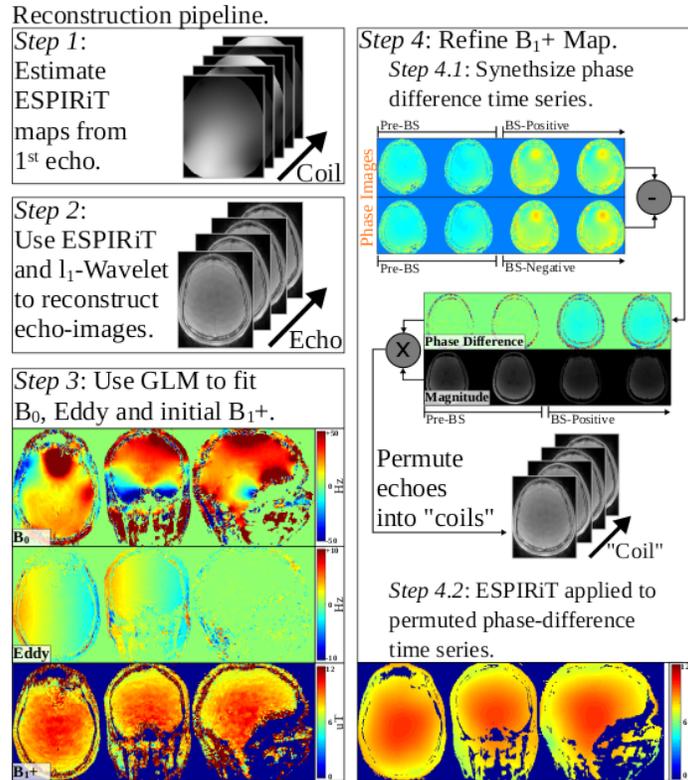


Figure 6-2: First, ESPIRiT is estimated from the ACS of the first echo. Next, a regularized linear inverse reconstruction with the wavelet prior estimates the echo images from the highly accelerated data. While artifacts are still present, they are incoherent over time. GLM robustly recovers  $B_0$ , but not  $B_1^+$ . Next, the phase of reconstructed echoes from  $BS^-$  is subtracted from the phase of the reconstructed echoes from  $BS^+$ . The resulting time-series is then reshaped into “virtual coils” and passed into ESPIRiT, which recovers the expected smooth underlying  $B_1^+$ .

## 6.4: Results

Figure (6-3) presents results from a selected undersampled case that achieves good trade-off with respect to speed versus recovered map quality. This constitutes an 11 second PhysiCal scan at over  $48\times$  acceleration across 12 echoes. Difference maps with respect to gold standard demonstrate high quality of reconstruction that accurately captures high frequency spatial variations in the  $B_0$  and  $B_1^+$  maps.

Figure (6-4) shows the results of the preliminary application of the method. The  $B_1^+$  obtained from the 11 second accelerated case is successful at mitigating striping artifacts in EPI-gSlider, and the  $B_0$  is successful at performing post-processing distortion correction to match the EPI image to a distortion-free structural image. However, it cannot resolve signal in voxel pile-up areas, which would require more data (for example, from a 2-shots blip-up and blip-down EPI)

## 6.5: Discussion

A rapid multi-parametric calibration scan termed PhysiCal is proposed and demonstrated to capture accurate high-resolution whole-brain  $B_0, B_1^+$  and coil sensitivity information in 11 seconds. This rapid acquisition is achieved through tailored undersampling and synergistic constrained eigenvalue reconstruction. Future work includes prospective undersampling implementation and evaluation at ultra-high field with along with utility in pTx calibration. It is expected that the increased variations in these applications will still be captured well by the eigenvalue approach given that ESPIRiT successfully estimates  $B_1^-$  for local

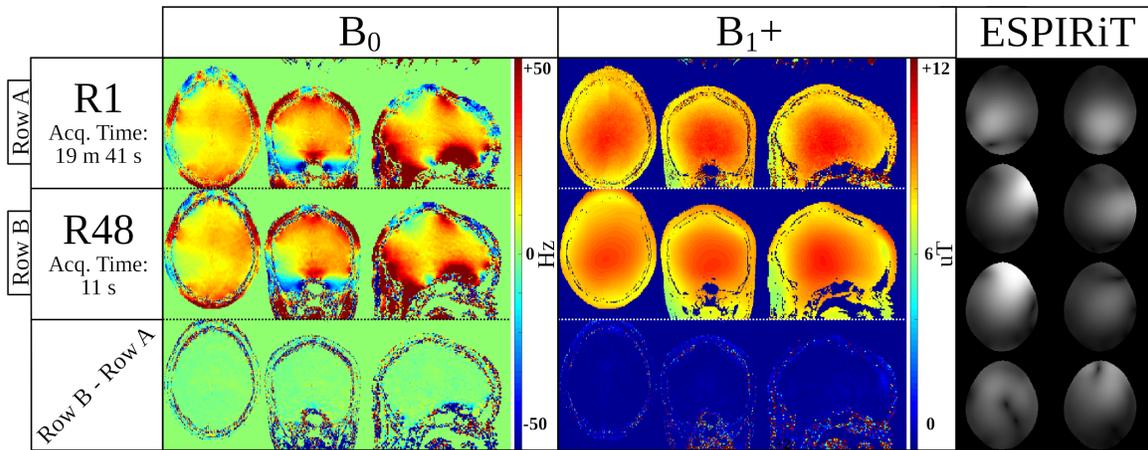


Figure 6-3: This figure compares the accelerated parametric maps (R48) to the gold standard (R1). By designing tailored under-sampling and using synergistic constrained and eigenvalue reconstruction, accurate and high resolution  $B_0$ ,  $B_1^+$ , and coil sensitivity maps can be recovered in 11 seconds.

coil arrays. More advanced reconstruction approaches through low-rank and phase-constraints will also be explored to aid an even faster and robust calibration scan. Additionally, the efficacy of PhysiCal will be verified on other applications including spatio-temporal encoding, such as like EPTI [102] and MRF [61, 15]

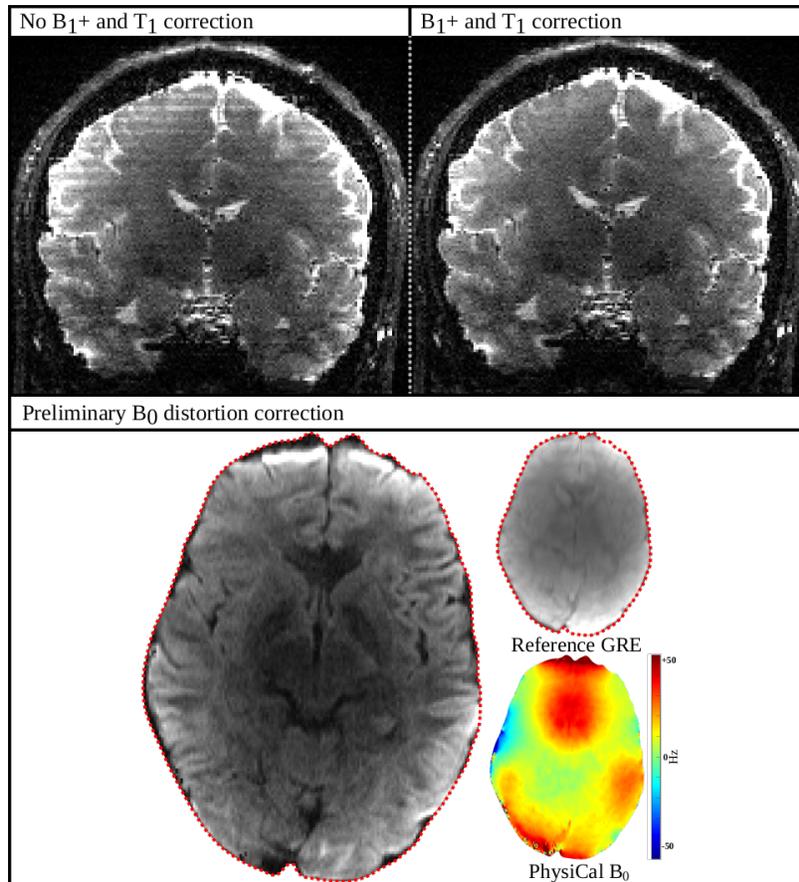


Figure 6-4: Preliminary applications of PhysiCal  $B_0$ ,  $B_1^+$  and coil sensitivity maps are recovered from a simulated 11-second PhysiCal scan.  $B_1^+$  is able to mitigate striping artifacts in EPI-gSlider acquisition.  $B_0$  is successful at performing post-processing distortion correction to un-distort the EPI image to match a distortion-free structural image. However, as expected, it is not able to resolve signal in the voxel pile-up areas, which would require more data (for example, from a 2-shots blipped-up and down EPI).



## 7: Conclusions

Spatio-Temporal MRI is a rapidly growing area of research due to the wealth of information it provides in a single examination. The major barrier to its general adoption is the long acquisition and reconstruction times. This thesis presented work done to improve the feasibility of such methods, with some future directions suggested below.

An interesting continuation of the polynomial preconditioner project is to theoretically study its effect on the Restricted Isometry Property and Robust Null Space Property [36]. On the practical side, approximating the polynomial preconditioner, say with a low-rank matrix, is expected to provide further practical speed up to the iterative convergence, particularly when the utilized proximal operator is relatively expensive to evaluate.

The Deli-CS framework utilized machine learning to jump-start an iterative algorithm for efficient reconstruction of TGAS-SPI-MRF data [15] in clinical setting. Future work would be to study how to effectively utilize multiple GPU devices in parallel for further reduction in reconstruction time, with the lofty goal of an approximately 1 minute reconstruction of 1 minute TGAS-SPI-MRF data so that the reconstruction would finish before the subsequent scan in the clinical MR exam is completed.

A learned regularization approach such as unrolled methods is expected to even further shorten the acquisition time of the same [42]. This would require the exploration of memory efficient training procedures given the large size of the  $A$  matrix.

The immediate next steps for the works presented in this thesis are to integrate and optimize the Physical acquisition, the TGAS-SPI-MRF acquisition and the Deli-CS reconstruction into a single, efficient pipeline so as to enable a rapid 1 mm isotropic resolution whole brain study in under 5 minutes (including reconstruction time) with BART [96] implementation.

On MRI acquisition, a metric to predict reconstruction quality in a manner that allows for sampling pattern design is needed. Currently, the trade-off between the least squares term of (1.1) and the prior  $g$  is determined qualitatively with no rigorous means of quantitative verification. This would be particularly useful for spatio-temporal MRI with subspace reconstruction, where a sampling pattern can be designed to better condition specific coefficients.

On MRI reconstruction in general, there is a need for a deterministic and rigorous regularization parameter selection process. Currently, this is done through trial-and-error with implicit factors like early stopping for regularization not rigorously quantified. This makes deploying such methods challenging as reconstruction quality in MRI is sensitive to chosen prior, regularization value and number of iterations. Ideally, an easy-to-use data driven training/ validation/ testing experiment framework that enables programmatic selection of hyper-parameters along with appropriate statistics to verify the choices will significantly reduce the manual effort required to tune a reconstruction while improving repeatability of the same. This can then be further augmented with techniques like [106].

With the advent of deep learning, there exists powerful cloud-based resources that can be leveraged for faster MRI reconstruction. This is a promising avenue for near real time processing with an interesting problem being how to best distribute the reconstruction. Similarly, leverage cloud compute requires addressing the challenge of compressing the acquired data so that it can be efficiently and privately transmitted to the servers.



# Bibliography

- [1] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 2018.
- [2] Shanshan Bao, Jonathan I Tamir, Jeffrey L Young, Umar Tariq, Martin Uecker, Peng Lai, Weitian Chen, Michael Lustig, and Shreyas S Vasanawala. Fast Comprehensive Single-sequence Four-dimensional Pediatric Knee MRI with T2 Shuffling. *Journal of Magnetic Resonance Imaging*, 2017.
- [3] Corey Baron, Nicholas Dwork, John M Pauly, and Dwight Nishimura. Rapid compressed sensing reconstruction of 3D non-Cartesian MRI. *Magnetic Resonance in Medicine*, 2018.
- [4] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2009.
- [5] Noam Ben-Eliezer, Daniel K Sodickson, and Kai Tobias Block. Rapid and Accurate T2 Mapping from Multi-spin-echo Data using Bloch-simulation-based Reconstruction. *Magnetic Resonance in Medicine*, 2015.
- [6] B Bilgic, H Ye, LL Wald, and K Setsompop. Optimized CS-wave Imaging with Tailored Sampling and Efficient Reconstruction. *Proceedings of the International Society of Magnetic Resonance in Medicine, Singapore*, 2016.
- [7] Berkin Bilgic, Borjan A Gagoski, Stephen F Cauley, Audrey P Fan, Jonathan R Polimeni, P Ellen Grant, Lawrence L Wald, and Kawin Setsompop. Wave-CAIPI for Highly Accelerated 3D Imaging. *Magnetic Resonance in Medicine*, 2015.
- [8] Berkin Bilgic, Tae Hyung Kim, Congyu Liao, Mary Kate Manhard, Lawrence L. Wald, Justin P. Haldar, and Kawin Setsompop. Improving parallel imaging by jointly reconstructing multi-contrast data. *Magnetic Resonance in Medicine*, 2018.
- [9] Kai Tobias Block, Martin Uecker, and Jens Frahm. Undersampled Radial MRI with Multiple Coils. Iterative Image Reconstruction Using a Total Variation Constraint. *Magnetic Resonance in Medicine*, 2007.
- [10] Moritz Blumenthal, Guanxiong Luo, Martin Schilling, H Christian M Holme, and Martin Uecker. Deep, deep learning with bart. *arXiv preprint arXiv:2202.14005*, 2022.
- [11] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [12] Felix A Breuer, Martin Blaimer, Robin M Heidemann, Matthias F Mueller, Mark A Griswold, and Peter M Jakob. Controlled Aliasing in Parallel Imaging Results in Higher Acceleration (CAIPIRINHA) for Multi-slice Imaging. *Magnetic Resonance in Medicine*, 2005.
- [13] Felix A Breuer, Martin Blaimer, Matthias F Mueller, Nicole Seiberlich, Robin M Heidemann, Mark A Griswold, and Peter M Jakob. Controlled Aliasing in Volumetric Parallel Imaging (2D CAIPIRINHA). *Magnetic Resonance in Medicine*, 2006.
- [14] Peng Cao, Xucheng Zhu, Shuyu Tang, Andrew Leynes, Angela Jakary, and Peder EZ Larson. Shuffled Magnetization-prepared Multicontrast Rapid Gradient-echo Imaging. *Magnetic Resonance in Medicine*, 2018.

- [15] Xiaozhi Cao, Congyu Liao, Siddharth Srinivasan Iyer, Zhixing Wang, Zihan Zhou, Erpeng Dai, Gilad Liberman, Zijing Dong, Ting Gong, Hongjian He, Jianhui Zhong, Berkin Bilgic, and Kawin Setsompop. Optimized multi-axis spiral projection mr fingerprinting with subspace reconstruction for rapid whole-brain high-isotropic-resolution quantitative imaging. *Magnetic Resonance in Medicine*, 2022.
- [16] Xiaozhi Cao, Huihui Ye, Congyu Liao, Qing Li, Hongjian He, and Jianhui Zhong. Fast 3D brain MR Fingerprinting based on Multi-Axis Spiral Projection Trajectory. *Magnetic Resonance in Medicine*, 2019.
- [17] Stephen F Cauley, Kawin Setsompop, Berkin Bilgic, Himanshu Bhat, Borjan Gagoski, and Lawrence L Wald. Autocalibrated Wave-CAIPI Reconstruction: Joint Optimization of k-space Trajectory and Parallel Imaging Reconstruction. *Magnetic Resonance in Medicine*, 2017.
- [18] Antonin Chambolle and Ch Dossal. On the Convergence of the Iterates of the “Fast Iterative Shrinkage/Thresholding Algorithm”. *Journal of Optimization Theory and Applications*, 2015.
- [19] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 2011.
- [20] Chen Chen, Lei He, Hongsheng Li, and Junzhou Huang. Fast iteratively reweighted least squares algorithms for analysis-based sparse reconstruction. *Medical Image Analysis*, 2018.
- [21] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Review*, 2001.
- [22] Nadège Corbin, Julio Acosta-Cabronero, Shaihan J Malik, and Martina F Callaghan. Robust 3d bloch-siegert based mapping using multi-echo general linear modeling. *Magnetic Resonance in Medicine*, 2019.
- [23] Gastão Cruz, Olivier Jaubert, Haikun Qi, Aurélien Bustin, Giorgia Milotta, Torben Schneider, Peter Koken, Mariya Doneva, René M Botnar, and Claudia Prieto. 3D Free-Breathing Cardiac Magnetic Resonance Fingerprinting. *Nuclear Magnetic Resonance in Biomedicine*, 2020.
- [24] Gastao Cruz, Olivier Jaubert, Torben Schneider, Rene M Botnar, and Claudia Prieto. Rigid Motion-Corrected Magnetic Resonance Fingerprinting. *Magnetic Resonance in Medicine*, 2019.
- [25] Mohammad Zalbagi Darestani, Akshay S Chaudhari, and Reinhard Heckel. Measuring robustness in deep learning based compressive sensing. *Proceedings of Machine Learning Research*, 2021.
- [26] Mohammad Zalbagi Darestani and Reinhard Heckel. Accelerated mri with un-trained neural networks. *IEEE Transactions on Computational Imaging*, 2021.
- [27] Mohammad Zalbagi Darestani, Jiayu Liu, and Reinhard Heckel. Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. *arXiv preprint arXiv:2204.07204*, 2022.
- [28] Ritobrato Datta, Micky K. Bacchus, Dushyant Kumar, Mark A. Elliott, Aditya Rao, Sudipto Dolui, Ravinder Reddy, Brenda L. Banwell, and Manojkumar Saranathan. Fast automatic segmentation of thalamic nuclei from MP2RAGE acquisition at 7 Tesla. *Magnetic Resonance in Medicine*, 2021.
- [29] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 2010.
- [30] Arjun D Desai, Beliz Gunel, Batu M Ozturkler, Harris Beg, Shreyas Vasanawala, Brian A Hargreaves, Christopher Ré, John M Pauly, and Akshay S Chaudhari. Vortex: Physics-driven data augmentations for consistency training for robust accelerated mri reconstruction. *arXiv preprint arXiv:2111.02549*, 2021.

- [31] Arjun D Desai, Batu M Ozturkler, Christopher M Sandino, Shreyas Vasanaawala, Brian A Hargreaves, Christopher M Re, John M Pauly, and Akshay S Chaudhari. Noise2recon: A semi-supervised framework for joint mri reconstruction and denoising. *arXiv preprint arXiv:2110.00075*, 2021.
- [32] Zijng Dong, Fuyixue Wang, Timothy G. Reese, Berkin Bilgic, and Kawin Setsompop. Echo planar time-resolved imaging with subspace reconstruction and optimized spatiotemporal encoding. *Magnetic Resonance in Medicine*, 2020.
- [33] William Falcon. Pytorch lightning. <https://github.com/PyTorchLightning/pytorch-lightning>, 2019.
- [34] Jeffrey A Fessler. Optimization Methods for Magnetic Resonance Image Reconstruction: Key Models and Optimization Algorithms. *IEEE Signal Processing Magazine*, 2020.
- [35] Jeffrey A Fessler, Sangwoo Lee, Valur T Olafsson, Hugo R Shi, and Douglas C Noll. Toeplitz-based Iterative Image Reconstruction for MRI with Correction for Magnetic Field Inhomogeneity. *IEEE Transactions on Signal Processing*, 2005.
- [36] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.
- [37] Pedro A Gómez, Matteo Cencini, Mohammad Golbabaee, Rolf F Schulte, Carolin Pirkl, Izabela Horvath, Giada Fallo, Luca Peretti, Michela Tosetti, Bjoern H Menze, et al. Rapid Three-Dimensional Multiparametric MRI with Quantitative Transient-State Imaging. *Scientific Reports*, 2020.
- [38] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 2002.
- [39] J. P. Haldar and D. Kim. OEDIPUS: An Experiment Design Framework for Sparsity-Constrained MRI. *IEEE Transactions on Medical Imaging*, 2019.
- [40] Justin P Haldar. Low-rank modeling of local  $k$ -space neighborhoods (loraks) for constrained mri. *IEEE Transactions on Medical Imaging*, 2014.
- [41] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic resonance in medicine*, 2018.
- [42] Kerstin Hammernik, Thomas Küstner, Burhaneddin Yaman, Zhengnan Huang, Daniel Rueckert, Florian Knoll, and Mehmet Akçakaya. Physics-Driven Deep Learning for Computational Magnetic Resonance Imaging. *arXiv preprint arXiv:2203.12215*, 2022.
- [43] Ken Hayami. Convergence of the Conjugate Gradient method on Singular Systems. *arXiv preprint arXiv:1809.00793*, 2018.
- [44] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [45] Chuan Huang, Ali Bilgin, Tomoe Barr, and Maria I. Altbach. T2 Relaxometry with Indirect Echo Compensation from Highly Undersampled Data. *Magnetic Resonance in Medicine*, 2013.
- [46] Olin G Johnson, Charles A Micchelli, and George Paul. Polynomial Preconditioners for Conjugate Gradient Calculations. *SIAM Journal on Numerical Analysis*, 1983.
- [47] Shmuel Kaniel. Estimates for Some Computational Techniques in Linear Algebra. *Mathematics of Computation*, 1966.
- [48] Steven Keckskemeti, Alexey Samsonov, Samuel A Hurley, Douglas C Dean, Aaron Field, and Andrew L Alexander. MPnRAGE: A Technique to Simultaneously Acquire Hundreds of Differently Contrasted MPnRAGE Images with Applications to Quantitative T1 Mapping. *Magnetic Resonance in Medicine*, 2016.

- [49] Daeun Kim, Stephen F Cauley, Krishna S. Nayak, Richard M. Leahy, and Justin P Haldar. Region-optimized virtual (ROVir) coils: Localization and/or suppression of spatial regions using sensor-domain beamforming. *Magnetic Resonance in Medicine*, 2021.
- [50] Daeun Kim, Jessica L Wisnowski, Christopher T Nguyen, and Justin P Haldar. Multidimensional correlation spectroscopic imaging of exponential decays: From theoretical principles to in vivo human applications. *NMR in Biomedicine*, 2020.
- [51] Kirsten Koolstra, Jeroen van Gemert, Peter Börnert, Andrew Webb, and Rob Remis. Accelerating compressed sensing in parallel imaging reconstructions using an efficient circulant preconditioner for cartesian trajectories. *Magnetic Resonance in Medicine*, 2019.
- [52] Thomas Küstner, Niccolo Fuin, Kerstin Hammernik, Aurelien Bustin, Haikun Qi, Reza Hajhosseiny, Pier Giorgio Masci, Radhouene Neji, Daniel Rueckert, René M Botnar, et al. CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Scientific Reports*, 2020.
- [53] Andreas Lesch, Matthias Schläegle, Martin Holler, Kristian Bredies, and Rudolf Stollberger. Ultrafast 3d bloch-siegert  $b_1^+$ -mapping using variational modeling. *Magnetic Resonance in Medicine*, 2019.
- [54] Andreas Johann Lesch, Andreas Petrovic, and Rudolf Stollberger. Robust implementation of 3D Bloch-Siegert  $B_1$  mapping. 2015.
- [55] Zhi-Pei Liang. Spatiotemporal Imaging with Partially Separable Functions. In *4<sup>th</sup> IEEE International Symposium on Biomedical Imaging*, 2007.
- [56] Congyu Liao, Jason Stockmann, Qiyuan Tian, Berkin Bilgic, Nicolas S Arango, Mary Kate Manhard, Susie Y Huang, William A Grissom, Lawrence L Wald, and Kawin Setsompop. High-fidelity, high-isotropic-resolution diffusion imaging through gslider acquisition with and  $t_1$  corrections and integrated  $\delta b_0$ /rx shim array. *Magnetic Resonance in Medicine*, 2020.
- [57] Congyu Liao Liao, Xiaozhi Cao, Siddharth Iyer, Zihan Zhou, Yunsong Liu, Justin Haldar, Mahmut Yurt, Ting Gong, Zhe Wu, Hongjian He, Jianhui Zhong Zhong, Adam Kerr, and Kawin Setsompop. Mesoscale myelin-water fraction and  $T_1/T_2/PD$  mapping through optimized 3D ViSta-MRF and stochastic reconstruction with preconditioning. *Proceedings of the International Society of Magnetic Resonance in Medicine*, 2022.
- [58] Bo Liu, Kevin King, Michael Steckner, Jun Xie, Jinhua Sheng, and Leslie Ying. Regularized Sensitivity Encoding (SENSE) Reconstruction Using Bregman Iterations. *Magnetic Resonance in Medicine*, 2009.
- [59] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 2007.
- [60] Michael Lustig, David L. Donoho, Juan M. Santos, and John M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 2008.
- [61] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L Sunshine, Jeffrey L Duerk, and Mark A Griswold. Magnetic Resonance Fingerprinting. *Nature*, 2013.
- [62] Merry Mani, Mathews Jacob, Vincent Magnotta, and Jianhui Zhong. Fast iterative algorithm for the reconstruction of multishot non-cartesian diffusion data. *Magnetic Resonance in Medicine*, 2015.
- [63] José P Marques, Tobias Kober, Gunnar Krueger, Wietske van der Zwaag, Pierre-François Van de Moortele, and Rolf Gruetter. MP2RAGE: A self bias-field corrected sequence for improved segmentation and  $T_1$ -mapping at high field. *Neuroimage*, 2010.

- [64] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, Amit Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. SymPy: Symbolic Computing in Python. *PeerJ Computer Science*, 2017.
- [65] Hisamoto Moriguchi and Jeffrey L Duerk. Bunched Phase Encoding (BPE): A New Fast Data Acquisition Method in MRI. *Magnetic Resonance in Medicine*, 2006.
- [66] Matthew J Muckley, Douglas C Noll, and Jeffrey A Fessler. Fast, Iterative Sub-Sampled Spiral Reconstruction via Circulant Majorizers. *Proceedings of the International Society of Magnetic Resonance in Medicine, Singapore*, 2016.
- [67] John P. Mugler III. Optimized Three-dimensional Fast-Spin-Echo MRI. *Journal of Magnetic Resonance Imaging*, 2014.
- [68] John P. Mugler III and James R. Brookeman. Three-dimensional Magnetization-Prepared Rapid Gradient-Echo Imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 1990.
- [69] Emilie Mussard, Tom Hilbert, Christoph Forman, Reto Meuli, Jean-Philippe Thiran, and Tobias Kober. Accelerated MP2RAGE imaging using Cartesian phyllotaxis readout and compressed sensing reconstruction. *Magnetic Resonance in Medicine*, 2020.
- [70] Yurii Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018.
- [71] Dwight G Nishimura. *Principles of Magnetic Resonance Imaging*. Stanford University, 2010.
- [72] Ola Norbeck, Tim Sprenger, Enrico Avventi, Henric Rydén, Annika Kits, Johan Berglund, and Stefan Skare. Optimizing 3D EPI for Rapid  $T_1$ -Weighted Imaging. *Magnetic Resonance in Medicine*, 2020.
- [73] Frank Ong, Shahab Amin, Shreyas Vasanawala, and Michael Lustig. mridata.org: An open archive for sharing MRI raw data. *Proceedings of the International Society of Magnetic Resonance in Medicine, Paris, France*, 2018.
- [74] Frank Ong and Michael Lustig. SigPy: A Python Package for High Performance Iterative Reconstruction. *Proceedings of the International Society of Magnetic Resonance in Medicine, Montréal, Canada*, 2019.
- [75] Frank Ong, Martin Uecker, and Michael Lustig. Accelerating Non-Cartesian MRI Reconstruction Convergence Using k-Space Preconditioning. *IEEE Transactions on Medical Imaging*, 2020.
- [76] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 2014.
- [77] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019.
- [78] Frederike H Petzschner, Irene P Ponce, Martin Blaimer, Peter M Jakob, and Felix A Breuer. Fast MR Parameter Mapping using k-t Principal Component Analysis. *Magnetic Resonance in Medicine*, 2011.
- [79] James G. Pipe and Padmanabhan Menon. Sampling density compensation in MRI: Rationale and an iterative numerical solution. *Magnetic Resonance in Medicine*, 1999.

- [80] Daniel Polak, Stephen Cauley, Berkin Bilgic, Enhao Gong, Peter Bachert, Elfar Adalsteinsson, and Kawin Setsompop. Joint multi-contrast variational network reconstruction (jVN) with application to rapid 2D and 3D imaging. *Magnetic Resonance in Medicine*, 2020.
- [81] Daniel Polak, Stephen Cauley, Susie Y Huang, Maria Gabriela Longo, John Conklin, Berkin Bilgic, Ned Ohringer, Esther Raithel, Peter Bachert, Lawrence L Wald, and Kawin Setsompop. Highly-accelerated Volumetric Brain Examination Using Optimized Wave-CAIPI Encoding. *Journal of Magnetic Resonance Imaging*, 2019.
- [82] Daniel Polak, Kawin Setsompop, Stephen F Cauley, Borjan A Gagoski, Himanshu Bhat, Florian Maier, Peter Bachert, Lawrence L Wald, and Berkin Bilgic. Wave-CAIPI for highly accelerated MP-RAGE imaging. *Magnetic Resonance in Medicine*, 2018.
- [83] Jonathan R Polimeni, Himanshu Bhat, Thomas Witzel, Thomas Benner, Thorsten Feiweier, Souheil J Inati, Ville Renvall, Keith Heberlein, and Lawrence L Wald. Reducing sensitivity losses due to respiration and motion in accelerated echo planar imaging by reordering the autocalibration data acquisition. *Magnetic Resonance in Medicine*, 2016.
- [84] Klaas P Pruessmann, Markus Weiger, Peter Börnert, and Peter Boesiger. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine*, 2001.
- [85] Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. SENSE: Sensitivity Encoding for Fast MRI. *Magnetic Resonance in Medicine*, 1999.
- [86] Sathish Ramani and Jeffrey A. Fessler. Parallel MR Image Reconstruction Using Augmented Lagrangian Methods. *IEEE Transactions on Medical Imaging*, 2011.
- [87] Laura I Sacolick, Florian Wiesinger, Ileana Hancu, and Mika W Vogel.  $b_1$  mapping by bloch-siegert shift. *Magnetic Resonance in Medicine*, 2010.
- [88] Christopher M Sandino, Peng Lai, Shreyas S Vasanawala, and Joseph Y Cheng. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magnetic Resonance in Medicine*, 2021.
- [89] Anne Marie Sawyer, Michael Lustig, Marcus Alley, Phdmartin Uecker, Patrick Virtue, Peng Lai, and Shreyas Vasanawala. Creation of fully sampled mr data repository for compressed sensing of the knee. *SMRT 22<sup>nd</sup> Annual Meeting, Salt Lake City, USA*, 2013.
- [90] Kawin Setsompop, Qiuyun Fan, Jason Stockmann, Berkin Bilgic, Susie Huang, Stephen F Cauley, Aapo Nummenmaa, Fuyixue Wang, Yogesh Rathi, Thomas Witzel, et al. High-resolution in vivo diffusion imaging of the human brain with generalized slice dithered enhanced resolution: Simultaneous multislice (gslider-sms). *Magnetic Resonance in Medicine*, 2018.
- [91] Jonathan Richard Shewchuk. *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*. School of Computer Science, Carnegie Mellon University, 1994.
- [92] Peter J Shin, Peder EZ Larson, Michael A Ohliger, Michael Elad, John M Pauly, Daniel B Vigneron, and Michael Lustig. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magnetic Resonance in Medicine*, 2014.
- [93] Jonathan I. Tamir, Valentina Taviani, Marcus T. Alley, Becki C. Perkins, Lori Hart, Kendall O'Brien, Fidaa Wishah, Jesse K Sandberg, Michael J. Anderson, Javier S. Turek, Theodore L. Willke, Michael Lustig, and Shreyas S. Vasanawala. Targeted rapid knee MRI exam using  $T_2$ -Shuffling. *Journal of Magnetic Resonance Imaging*, 2019.
- [94] Jonathan I Tamir, Martin Uecker, Weitian Chen, Peng Lai, Marcus T Alley, Shreyas S Vasanawala, and Michael Lustig.  $T_2$ -Shuffling: Sharp, Multicontrast, Volumetric Fast Spin-Echo Imaging. *Magnetic Resonance in Medicine*, 2017.

- [95] Martin Uecker, Thorsten Hohage, Kai Tobias Block, and Jens Frahm. Image reconstruction by regularized nonlinear inversion - joint estimation of coil sensitivities and image content. *Magnetic Resonance in Medicine*, 2008.
- [96] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanaawala, and Michael Lustig. ESPIRiT - An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE Meets GRAPPA. *Magnetic Resonance in Medicine*, 2014.
- [97] Martin Uecker, Frank Ong, Jonathan I Tamir, Dara Bahri, Patrick Virtue, Joseph Y Cheng, Tao Zhang, and Michael Lustig. Berkeley Advanced Reconstruction Toolbox. *Proceedings of the International Society of Magnetic Resonance in Medicine, Toronto, Canada*, 2015.
- [98] Abraham Van der Sluis and Henk A van der Vorst. The Rate of Convergence of Conjugate Gradients. *Numerische Mathematik*, 1986.
- [99] Julia V Velikina, Andrew L Alexander, and Alexey Samsonov. Accelerating MR Parameter Mapping using Sparsity-promoting Regularization in Parametric Dimension. *Magnetic Resonance in Medicine*, 2103.
- [100] Julia V Velikina and Alexey A Samsonov. Reconstruction of Dynamic Image Series from Undersampled MRI Data using Data-driven Model Consistency Condition (MOCCO). *Magnetic Resonance in Medicine*, 2015.
- [101] FTAW Wajer and KP Pruessmann. Major speedup of reconstruction for sensitivity encoding with arbitrary trajectories. *Proceedings of the International Society of Magnetic Resonance in Medicine, Toronto, Canada*, 2001.
- [102] Fuyixue Wang, Zijing Dong, Timothy G Reese, Berkin Bilgic, Mary Katherine Manhard, Jingyuan Chen, Jonathan R Polimeni, Lawrence L Wald, and Kawin Setsompop. Echo Planar Time-Resolved Imaging (EPTI). *Magnetic Resonance in Medicine*, 2019.
- [103] Xiaoqing Wang, Zhengguo Tan, Nick Scholand, Volkert Roeloffs, and Martin Uecker. Physics-based reconstruction methods for magnetic resonance imaging. *Philosophical Transactions of the Royal Society*, 2021.
- [104] Matthias Weigel. Extended Phase Graphs: Dephasing, RF Pulses, and Echoes - Pure and Simple. *Journal of Magnetic Resonance Imaging*, 2015.
- [105] Daniel S. Weller, Sathish Ramani, and Jeffrey A. Fessler. Augmented Lagrangian with Variable Splitting for Faster Non-Cartesian  $L_1$ -SPIRiT MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 2014.
- [106] Daniel S Weller, Sathish Ramani, Jon-Fredrik Nielsen, and Jeffrey A Fessler. Monte Carlo SURE-based parameter selection for parallel Magnetic Resonance Imaging reconstruction. *Magnetic Resonance in Medicine*, 2014.
- [107] Zheng Xu, Sheng Wang, Yeqing Li, Feiyun Zhu, and Junzhou Huang. PRIM: An Efficient Preconditioning Iterative Reweighted Least Squares Method for Parallel Brain MRI Reconstruction. *Neuroinformatics*, 2018.
- [108] Huihui Ye, Dan Ma, Yun Jiang, Stephen F Cauley, Yiping Du, Lawrence L Wald, Mark A Griswold, and Kawin Setsompop. Accelerating magnetic resonance fingerprinting (MRF) using t-blipped simultaneous multislice (SMS) acquisition. *Magnetic Resonance in Medicine*, 2016.
- [109] Leslie Ying and Jinhua Sheng. Joint Image Reconstruction and Sensitivity Estimation in SENSE (JSENSE). *Magnetic Resonance in Medicine*, 2007.
- [110] Benjamin Zahneisen, Murat Aksoy, Julian Maclaren, Christian Wuerstin, and Roland Bammer. Extended hybrid-space SENSE for EPI: Off-resonance and eddy current corrected joint interleaved blip-up/down reconstruction. *NeuroImage*, 2017.

- [111] Bo Zhao, Wenmiao Lu, T Kevin Hitchens, Fan Lam, Chien Ho, and Zhi-Pei Liang. Accelerated MR Parameter Mapping with Low-rank and Sparsity Constraints. *Magnetic Resonance in Medicine*, 2015.
- [112] Bo Zhao, Kawin Setsompop, Elfar Adalsteinsson, Borjan Gagoski, Huihui Ye, Dan Ma, Yun Jiang, P Ellen Grant, Mark A. Griswold, and Lawrence L. Wald. Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling. *Magnetic Resonance in Medicine*, 2018.
- [113] Md. Zulfiqar Ali Bhotto, M. Omair Ahmad, and M. N. S. Swamy. An Improved Fast Iterative Shrinkage Thresholding Algorithm for Image Deblurring. *SIAM Journal on Imaging Sciences*, 2015.