

# **An Adaptive Signal Processing Approach to Dynamic Magnetic Resonance Imaging**

A Thesis Presented

by

William Scott Hoge

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the field of

Communications and Signal Processing

Northeastern University  
Boston, Massachusetts

May 2001


**NORTHEASTERN UNIVERSITY**  
**Graduate School of Engineering**

Thesis Title: An Adaptive Signal Processing Approach to Dynamic Magnetic Resonance Imaging

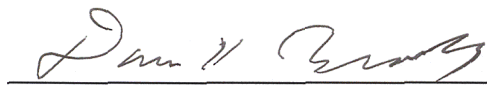
Author: William Scott Hoge, Jr.

Department: Electrical and Computer Engineering

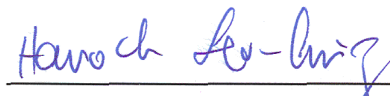
Approved for Thesis Requirement of the Doctor of Philosophy Degree

  
\_\_\_\_\_  
Thesis Advisor, Eric L. Miller


5/21/2001  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Thesis Committee Member, Dana H. Brooks


5/21/2002  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Thesis Committee Member, Hanoch Lev-Ari


5/21/2001  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Thesis Committee Member, Lawrence P. Panych  
Harvard Medical School and  
Brigham and Women's Hospital

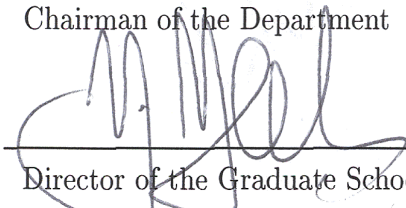
5/22/2001  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Thesis Committee Member, W. Clement Karl  
Boston University

5/23/2001  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Chairman of the Department

5/22/01  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Director of the Graduate School

5/23/01  
\_\_\_\_\_  
Date

# ABSTRACT

Magnetic resonance imaging (MRI) is a powerful non-invasive imaging tool that has found extensive use in medical diagnostic procedures. Dynamic MRI refers to the acquisition of multiple images in order to observe changes in tissue structure over time. Clinical applications include the observation of the early flow of contrast agent to detect tumors and real time monitoring of surgical interventions and thermal treatments.

The primary goal of our research is to reduce the acquisition time of dynamic MRI sequences through the application of signal processing concepts. These concepts include adaptive filtering techniques, system subspace identification, and subspace tracking. Presented in this thesis are methods to find estimates of the true sequence images from a limited amount of acquired data using optimization of multiparameter function techniques. The methods build on the linear MRI system response model first proposed by Panych and Zientara.

Three new methods related to dynamic MRI are presented. First, because medically significant changes are typically limited to a small region of interest (ROI), a static ROI estimation problem is presented along with a numerical solution algorithm. This static problem has parallels to matrix completion problems in the field of linear algebra. Second, a general adaptive image estimation framework for dynamic MRI is described. Analysis shows that most previous low-order methods are special cases of this general framework. Third, two methods are presented for identifying suitable MR data acquisition inputs to use with the adaptive estimation framework: one relies on a conjugate gradient algorithm constrained to the Stiefel manifold; the second relies on linear prediction. The combination of the adaptive estimation framework and dynamic input identification methods provide a mechanism to efficiently track changes in an image slice, potentially enabling significant acquisition time savings in a clinical setting.

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors for their time and expert guidance in the preparation of this manuscript. Special thanks goes to Dr. Eric L. Miller for providing the opportunity for this research. I owe much to Dr. Dana H. Brooks and Dr. Hanoch Lev-Ari as well for their bountiful support and fruitful conversations. Thanks go as well to Dr. Lawrence P. Panych for motivating such an interesting problem and providing laboratory time and data.

A very special thank you goes to my family, especially my parents and my wife Melissa. I am deeply indebted to them for providing me encouragement and support every step of the way.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Fundamental Physics of MRI</b>	<b>6</b>
2.1	Dynamics from a modern physics perspective . . . . .	7
2.2	Dynamics from a classical physics perspective . . . . .	9
2.2.1	Precession . . . . .	9
2.2.2	Relaxation . . . . .	12
2.2.3	The Bloch equations . . . . .	14
2.2.4	An example of spin manipulation: simple (Hahn) spin echo . . . . .	15
2.2.5	Classical dynamics summary . . . . .	16
2.3	Acquisition of an image . . . . .	17
2.3.1	Signal detection . . . . .	17
2.3.2	Gradient fields, spin density, and $k$ -space . . . . .	18
2.3.3	Selective excitation . . . . .	19
2.3.4	2-D Fourier imaging . . . . .	20
2.4	Summary . . . . .	22
<b>3</b>	<b>Image acquisition via low order encoding</b>	<b>23</b>
3.1	Fourier based methods . . . . .	24
3.1.1	Fourier Keyhole . . . . .	24
3.1.2	RIGR . . . . .	25
3.2	A linear system model for non-Fourier based methods . . . . .	26
3.2.1	SVD encoding method . . . . .	30
3.2.2	The relationship between spatial and $k$ -space representations of an image . . . . .	30
3.3	Useful error measures . . . . .	32
3.3.1	Measuring distance between images and estimates . . . . .	33
3.3.2	Measuring distance between subspaces . . . . .	33
3.4	Summary . . . . .	34
<b>4</b>	<b>Efficient region of interest acquisition</b>	<b>35</b>
4.1	Problem formulation . . . . .	35
4.2	Minimal order problem . . . . .	36
4.2.1	Rectangular ROI, arbitrary error threshold . . . . .	37
4.2.2	Arbitrarily specified ROI, zero error . . . . .	39
4.3	Minimal error, fixed order problem . . . . .	41
4.3.1	CCD algorithm . . . . .	42
4.3.2	CCD algorithm initialization . . . . .	43
4.3.3	Choice of approximation order . . . . .	44
4.4	Examples . . . . .	45
4.4.1	Simulation results . . . . .	45

4.4.2	Laboratory results . . . . .	50
4.5	Summary of the static problem . . . . .	52
<b>5</b>	<b>Adaptive modeling of the dynamic MRI process</b>	<b>57</b>
5.1	Construction of the image estimate . . . . .	58
5.2	Input vector identification . . . . .	62
5.2.1	The subspace trap . . . . .	63
5.2.2	Escaping the subspace trap I: CG-St . . . . .	66
5.2.3	Escaping the subspace trap II: Image prediction . . . . .	79
5.3	Method comparison examples . . . . .	84
5.4	Summary of the dynamic problem . . . . .	108
<b>6</b>	<b>Conclusions and future research</b>	<b>118</b>
6.1	Open static problem questions . . . . .	119
6.2	Open dynamic problem questions . . . . .	122
<b>A</b>	<b>Analytic Details</b>	<b>125</b>
A.1	Linear Algebra Nomenclature . . . . .	125
A.2	Derivatives of complex valued matrix functions . . . . .	129
A.3	Efficient solution of vectorized systems . . . . .	130
A.4	Index of symbols . . . . .	131
	<b>Bibliography</b>	<b>135</b>

# List of Figures

2.1	Magnetization vector and transverse plane projection . . . . .	12
2.2	2-D Fourier Acquisition Timing Diagram . . . . .	21
2.3	Using magnetic field gradients to scan k-space . . . . .	21
4.1	Original MR Image and ROI for static simulation example . . . . .	46
4.2	Permuted selection matrix for Figure 4.1 and geometric determination of $r_u$ . . . . .	46
4.3	Relative error comparison of SVD, LoF, and CCD solutions for Figure 4.1 ROI. . . . .	47
4.4	Comparison of order $r = 10$ ROI reconstructions . . . . .	48
4.5	Pixel value difference comparisons of order $r = 10$ ROI reconstructions . . . . .	49
4.6	Comparison of order $r = 25$ ROI reconstructions . . . . .	49
4.7	Pixel value difference comparisons of order $r = 25$ ROI reconstructions . . . . .	50
4.8	Original phantom image for static laboratory example . . . . .	51
4.9	Covering ROI for lab experiments . . . . .	52
4.10	ROI reconstruction for $X = U$ . . . . .	53
4.11	ROI reconstruction for $X = U\Sigma^{1/2}$ . . . . .	53
4.12	ROI reconstruction for $X = U\Sigma$ . . . . .	54
4.13	Interior ROI for lab result example . . . . .	54
4.14	Reconstruction of interior ROI images . . . . .	55
5.1	Comparison between the steepest descent and conjugate gradient methods . . . . .	72
5.2	Contrast change example reference image showing ROI . . . . .	86
5.3	Original synthetic contrast change sequence . . . . .	88
5.4	Relative error comparison of low-order acquisition methods: synthetic contrast change . . . . .	89
5.5	Simulated synthetic contrast change sequence acquisition using the Optimal method . . . . .	90
5.6	Simulated synthetic contrast change sequence acquisition using the Linear Predictor method: $\text{lp}(YX^H)$ . . . . .	91
5.7	Simulated synthetic contrast change sequence acquisition using the Linear Predictor method: $\text{lp}(\text{Aest})$ . . . . .	92
5.8	Simulated synthetic contrast change sequence acquisition using the CG-St method . . . . .	93
5.9	Simulated synthetic contrast change sequence acquisition using the Fourier Keyhole method . . . . .	94
5.10	Simulated synthetic contrast change sequence acquisition using the keyhole SVD method . . . . .	95
5.11	Simulated synthetic contrast change sequence acquisition using the RIGR method . . . . .	96
5.12	Low-order acquisition method comparison showing the relative error estimating actual contrast change MRI data . . . . .	97
5.13	Low-order acquisition method comparison showing the relative error for the rapid acquisition synthetic contrast change sequence . . . . .	99
5.14	Original synthetic grapefruit sequences . . . . .	101
5.15	Relative error comparison for basic synthetic grapefruit sequence . . . . .	102

5.16	Relative error comparison for synthetic grapefruit sequence with section expansion	103
5.17	Relative error comparison for synthetic grapefruit sequence with random jitter . .	104
5.18	Relative error comparison for synthetic grapefruit sequence with section expansion and random jitter . . . . .	105
5.19	Original image sequence for simulated grapefruit acquisition . . . . .	109
5.20	Relative error comparison for simulated grapefruit sequence acquisition . . . . .	109
5.21	Simulated grapefruit sequence acquisition using Optimal method . . . . .	110
5.22	Simulated grapefruit sequence acquisition using Linear Predictor method: $\text{lp}(YX^H)$	111
5.23	Simulated grapefruit sequence acquisition using Linear Predictor method: $\text{lp}(\text{Aest})$	112
5.24	Simulated grapefruit sequence acquisition using CG-St method . . . . .	113
5.25	Simulated grapefruit sequence acquisition using Fourier Keyhole method . . . . .	114
5.26	Simulated grapefruit sequence acquisition using keyhole SVD method . . . . .	115
5.27	Simulated grapefruit sequence acquisition using RIGR method . . . . .	116
6.1	Diamond shaped region of interest . . . . .	121



# List of Tables

2.1	Simple spin echo sequence . . . . .	16
3.1	Fourier keyhole dynamic sequence acquisition method . . . . .	25
3.2	RIGR dynamic sequence acquisition method . . . . .	27
3.3	SVD dynamic sequence acquisition method . . . . .	31
5.1	Image reconstruction method summary . . . . .	62
5.2	Efficient conjugate gradient algorithm . . . . .	73
5.3	Conjugate gradient for minimizing $F(X)$ on the Stiefel manifold . . . . .	79
5.4	Predetermined equations for image prediction from uniformly sampled image estimates	82
5.5	Table of synthetic test sequences and associated figures . . . . .	100
5.6	Methods and associated figures used in simulated acquisition using actual MRI data	107
6.1	Error comparison between CCD and $X = L \in St(n, r)$ methods for 100 random matrices . . . . .	121

# Chapter 1

## Introduction

Medical imaging technology has seen dramatic advances over recent years. One method that has become a very powerful tool for imaging soft tissue is magnetic resonance imaging (MRI). MRI has found extensive use in a variety of medical diagnostic procedures because it provides high contrast images of internal tissue structure through non-invasive means. According to [45], MRI has become the imaging modality of choice for diagnostic studies of the head, spine, and joints.

The term *dynamic MRI* refers to acquisition of a sequence of images to monitor changes in tissue structure over time [34]. Clinical applications where dynamic MRI is of interest include the observation of the early flow of contrast agent to detect tumors [42, 44], real time monitoring of surgical interventions or thermal treatments [22], and cardiac imaging [46]. Because of limits in the data acquisition rate, there is a trade-off in each of these cases between temporal resolution, spatial resolution, volume coverage and signal-to-noise ratio. For example, the ability to image cardiac activity in real time comes at the expense of limited volume coverage and low spatial resolution [23]. Thus, there is a need for optimized data acquisition that allows faster image sequence acquisition with less data.

Traditional MRI acquisition techniques use a series of magnetic field gradients and radio-frequency (rf) pulses to encode the position of different particles within a tissue volume. These

excitation sequences are used to scan a volume in a sequence of slices, typically by direct sampling of the two-dimensional spatial Fourier domain, or  $k$ -space, of the slice. An inverse Fourier transform is then used to reconstruct images of the tissue composition within each slice. A review of these traditional imaging techniques is provided in Chapter 2. Good reviews from a signal processing perspective are also available in [45] and [26].

The physical dynamics of MR imaging constrain the image acquisition time. Typically, one line in  $k$ -space is sampled for each input excitation sequence. For single  $k$ -space line sampling techniques, the required image acquisition time is proportional to the number of lines sampled in  $k$ -space, or equivalently, the number of excitations used. One approach to reduce the acquisition time of a single image is to lower the number of excitations employed and obtain a *low order* representation of the underlying image. Thus the problem of reducing acquisition time is equivalent to designing both new image reconstruction models and excitation sequences to reconstruct estimates of the images.

Multi-line sampling techniques are also available, but these typically require enhanced hardware to implement. For example, echo-planar imaging (EPI) samples a cyclic raster line through  $k$ -space, but requires quickly switching a strong magnetic gradient field [5, p. 152]. A second example, SENSE [38], uses a phased array of receiver coils to rapidly sample  $k$ -space. Both of these methods represent a hardware solution. In contrast, the low-order methods discussed in this thesis are a *software* approach to image acquisition. The two approaches are complimentary [35], thus the discussion here is limited to single  $k$ -space line sampling methods.

In this work the main approach to reducing the image acquisition time is through the application of signal processing concepts. We approach the dynamic MRI problem from two perspectives. One comes from the observation that in typical dynamic MRI sequences, the medically significant changes occur in a limited region of interest (ROI). Imaging tissue outside the ROI consumes both time and resources, and yet provides only extraneous information. If the ROI could be adequately reconstructed using a relatively small number of excitations, then the time to acquire

the ROI would be correspondingly reduced. Thus, the problem of identifying appropriate excitation sequences and reconstruction vectors to represent an arbitrarily shaped region of interest in a given image is first examined. As in previous full image approaches, we assume a known prior image and use it to design appropriate image acquisition sequences. This *static problem* is quite similar to image representation [19] and matrix completion [8] problems. Second, all of the low-order acquisition methods previously proposed rely on the premise that future images in a sequence are not “significantly” different from previous images. That is, they draw on knowledge of a past history of full-order images to design low-order system excitation and image reconstruction strategies. We refer to this as the *dynamic problem*. The solutions to the dynamic problem presented later in this thesis draw from concepts such as adaptive filtering techniques [14], system subspace identification [40], and subspace tracking [41].

Both the static and dynamic problems concern finding methods that identify low-order estimates of the true sequence images. These methods strive to achieve minimal error between the true images and the image estimates based on criteria described in Section 3.3 below. In both cases, the identification of appropriate estimates is achieved through the optimization of multiparameter functions. This optimization is approached both analytically and numerically using function gradient and gradient descent techniques.

The significant results of this work are the following. For the static problem, a numerical method is presented in Chapter 4 to efficiently represent an arbitrarily shaped region of interest in a static image. This method is a significant addition to the body of matrix completion problem solutions. For unlike traditional matrix completion problems, this new method does not impose any presumed structure on the matrix to guide the solution method. However, as Section 4.4.2 shows, the utility of this method for acquiring MRI images is somewhat limited. The presence of noise in the image acquisition process severely corrupts the ability of this method to provide high quality estimates of the ROI. Possible methods to repair this shortcoming of the algorithm are discussed in Chapter 6.

For the dynamic problem, three significant results are presented in this thesis. As discussed in Chapter 5, the dynamic problem can be segmented into two related problems: *Image Estimation* and *Input Identification*. Building on the linear system response model first developed by Panych and Zientara [36], Section 5.1 presents a general adaptive framework for dynamic image estimation. Analysis of this framework shows that most of the previously proposed low-order acquisition methods are special cases of the general adaptive framework presented here. To complement this framework, two system input identification methods are presented in Section 5.2. One of the conclusions of the adaptive framework analysis is that orthonormal input vector sets are extremely beneficial. Thus the first method presented, CG-St, seeks to find an optimal set of inputs by constraining a minimization problem to the parameter space of orthonormal matrices. This approach provides new input vectors that are less biased towards previous inputs than previous methods allowed. However, even greater performance improvement is provided by a second input identification method,  $lp(\cdot)$ , which uses a linear predictor to determine new input vectors. Both the CG-St and  $lp(\cdot)$  methods outperform previously proposed low-order acquisition methods in a variety of synthetic scenarios and, more importantly, in dynamic sequence acquisition simulations using real MRI data.

The work presented in this thesis was directed by my thesis committee: Eric L. Miller, Dana H. Brooks, and Hanoch Lev-Ari, and was performed in collaboration with Lawrence P. Panych of the Radiology Department, Harvard Medical School (Brigham and Women’s Hospital, Boston). W. Clem Karl, Boston University, has also provided invaluable assistance with this project.

The structure of the document follows a path similar to the topics discussed above. First a review of the physics fundamental to the acquisition of magnetic resonance images is presented in Chapter 2. A brief overview of traditional Fourier based image acquisitions is also presented. Next, a review of the current “state of the low-order acquisition art” is given in Chapter 3. This section reviews in detail the linear system response model on which the new imaging methods presented

in this thesis are based. Chapter 4 presents the static problem, including both simulation and laboratory examples. Chapter 5 presents a general adaptive framework for the estimation of dynamic image sequences along with two input identification techniques. Chapter 6 closes the thesis with a discussion of avenues available for future research. A brief review of background topics needed in this thesis is included in the Appendix. This includes a review of linear algebra nomenclature and concepts in Appendix A.1 and a discussion on finding the derivatives of complex valued matrix functions in Appendix A.2.

Enjoy!

## Chapter 2

# The Fundamental Physics of MRI

Magnetic resonance imaging (MRI) was introduced to the world in 1973. With two short pages in the journal *Nature* [24], P. C. Lauterbur described how to discern the location and composition of different material through the application of electro-magnetic fields. The basic principle is to electro-magnetically encode the spatial location and composition of material to be imaged, scan the encoding, and reconstruct images from the recorded data. The strength of MRI is that images of soft tissue structure can be reconstructed through non-invasive means. A second advantage is that the imaging method is also non-destructive, since MRI relies on the ability of particles in a magnetic field to store and release energy rather than absorbing the energy as in X-ray imaging. This chapter seeks to describe the fundamental physical models of MRI imaging.

The MR imaging process can be modeled at a variety of levels, from low-level atomic interaction modeling to abstract system modeling. This chapter presents a wide spectrum of these models, provides a theoretical foundation for the imaging process, and gives some context for the advanced low-order imaging methods described in the remainder of the thesis. The topics presented include the quantum mechanical behavior of material (spin) that is manipulated in the imaging process, a classical dynamics description of an aggregate collection of atoms with spin, how spin is manipulated with electro-magnetic energy, and how this manipulation of spin can produce an image of

tissue structure through non-invasive means.

## 2.1 Dynamics from a modern physics perspective

Quantum mechanics models the workings of the atomic world. One of the findings of the past century was that the mechanical model of angular momentum from classical physics, i.e., the “spinning-top”, leads to contradictions with experimental results at the atomic particle level. For example, the experimentally observed magnetic moment associated with the angular momentum of an electron turns out to be twice as large as the classical model predicts. This inconsistency was resolved by Pauli through the introduction of *spin operators* [37].

*Spin* is the description of the intrinsic angular momentum observed in atomic particles that is distinct from orbital angular momentum. The observed angular momentum is a combination of the spin, a quantum physics modeling of the dynamics, and the orbital angular momentum derived from modeling the dynamics from a classical physics description. The term *spin* was chosen to emphasize the distinction. Dirac showed that the description of quantum spin is in fact a specific form of an abstract operator. These operators allow calculation of the spin quantum values,  $\vec{I}$ , algebraically. The range of  $\vec{I}$  is limited to a series of discrete values. In the presence of an external magnetic field these values are  $\{-\frac{1}{2}, +\frac{1}{2}\}$  for many of the nuclei typically used for MR imaging, such as the hydrogen atom.

From the theory of quantum mechanics, one can describe the relation between the spin angular momentum,  $\hbar\vec{I}$ , and the molecular magnetic moment,  $\vec{\mu}$ , via by the following relation

$$\vec{\mu} = \gamma\hbar\vec{I}$$

where  $\hbar$  is Planck’s constant divided by  $2\pi$ ,  $\vec{I}$  is a dimension-less angular momentum vector describing the intrinsic spin state, and  $\gamma$  is the gyrometric ratio which depends on the sign, size, and distribution of charge within the material. When placed in a magnetic field  $\vec{B}$ , these magnetic



moments will polarize with energy

$$E = -\vec{\mu} \cdot \vec{B}. \quad (2.1)$$

If the magnetic field is oriented along the  $z$ -axis, e.g.,  $\vec{B} = B_0 \hat{a}_z$ , then  $E = -\mu_z B_0$ . For nuclei with potential quantum states  $m = \{-\frac{1}{2}, +\frac{1}{2}\}$ , this implies that the potential energy states are

$$E_m = \{-\gamma \hbar B_0 / 2, +\gamma \hbar B_0 / 2\}. \quad (2.2)$$

If a quanta of irradiated energy of magnitude  $\gamma \hbar B_0$  is absorbed by the nuclei, the polarization of a particle will change to the higher energy state.

When a collection of spins are at thermal equilibrium, the spin state population density is dictated by Boltzmann statistics. That is, the probability of finding a particle in a specific spin state is proportional to  $\exp\{-E_m/kT\}$ . Here,  $k$  is the Boltzmann constant,  $E_m$  is the energy of particle at state  $m$ , and  $T$  is temperature. By averaging over all possible spin states, the aggregate magnetization is given by

$$M_0 = \rho \gamma \hbar \frac{\sum m e^{-E_m/kT}}{\sum e^{-E_m/kT}} \quad (2.3)$$

where  $\rho$  is the number of nuclei per unit volume and the summation is performed over all possible energy states. At room temperature,  $E_m \ll kT$  and the exponential terms may be approximated by  $(1 - E_m/kT)$ . In general, for nuclei with spin  $I$ , this allows the magnetization to be approximated as

$$M_0 \simeq \rho \gamma^2 \hbar^2 \left[ \frac{I(I+1)}{3kT} \right] B_0.$$

Continuing the example of one-half spin nuclei, i.e.,  $I = 1/2$  and  $m = \{+\frac{1}{2}, -\frac{1}{2}\}$ , the magnetization vector is

$$M_0 \simeq \rho \frac{\gamma^2 \hbar^2}{4kT} B_0. \quad (2.4)$$

The significance of this relation is that the magnetization depends primarily on the quantum spin states  $I$ , the applied magnetic field  $B_0$ , the temperature  $T$  of the system, and the distribution  $\rho$  of the spins through the volume. The remaining parameters are intrinsic constants. [1]

In summary, for the remainder of this thesis one need only be concerned with the following conclusions from the quantum mechanical description of the imaging dynamics. First, spin is an intrinsic property of matter, observable only at the atomic level. Second, each particle has a magnetic moment that is directly proportional to both the intrinsic spin and the composition of the particle, described by  $\gamma$ . Third, an aggregate collection of spins can be approximated by a bulk magnetization term. The magic of MRI is that through manipulation of these magnetic moments, one can non-invasively construct an image.

While quantum physics completely describes particle dynamics in a magnetic field, it is also cumbersome to describe a large collection of particles. Thus we move now to a classical physics perspective and examine the effect of magnetic fields on the bulk magnetization.

## 2.2 Dynamics from a classical physics perspective

This section reviews the behavior of the bulk magnetization in a magnetic field. The magnetization arises from the intrinsic angular momentum, or spin, of the atomic particles within a volume of tissue. By looking at the aggregate collection of the spins, the magnetization motion can be analyzed from a classical perspective.

### 2.2.1 Precession

As shown in Section 2.1, the motion of an ensemble of independent spin one-half nuclei in a magnetic field may be described in terms of the spin magnetization vector,  $\mathbf{M}$ . By definition, the magnetization is proportional to the angular momentum  $\mathbf{L}$

$$\mathbf{M} = \gamma \mathbf{L} \tag{2.5}$$

where the gyrometric ratio  $\gamma$  depends on the sign, size, and distribution of charge within the material. The torque acting on the magnetization in a magnetic field  $\mathbf{B}$  is given as

$$\text{Torque} = \frac{d\mathbf{L}}{dt} = \mathbf{M} \times \mathbf{B}. \quad (2.6)$$

Combining these two equations, one obtains

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B}. \quad (2.7)$$

When the magnetization is oriented parallel to  $\mathbf{B}$ ,  $d\mathbf{M}/dt = 0$ . This is considered the equilibrium state. When  $\mathbf{M}$  is not parallel to  $\mathbf{B}$ , the solution to Equation (2.7) when  $\mathbf{B}$  is a magnetic field of amplitude  $B_0$  corresponds to a precession of the magnetization about the field at the rate  $\omega_0 = \gamma B_0$ , the *Larmor frequency*. For reference, the static magnetic field,  $\mathbf{B}_0$ , is assumed to be oriented along the positive  $z$  axis,  $\hat{a}_z$ , for the remainder of the discussion. Precession is so common in MRI that it is useful to consider Equation (2.7) in a reference frame rotating about the  $z$ -axis at an angular frequency  $\omega$ . This “reference frame moment” is denoted  $\vec{\omega}$ . The velocity  $\mathbf{v}$  of a particle in this rotating frame can be described by

$$\mathbf{v} = \mathbf{v}_a + \vec{\omega} \times \mathbf{r} \quad (2.8)$$

where  $\mathbf{v}_a$  is the actual velocity in a fixed frame and the cross term represents the rotating frame translation to the fixed frame. If the magnetization is directed along  $\mathbf{r}$ , the rate of magnetization change can be described by

$$\frac{d\mathbf{M}}{dt} = \gamma \frac{d\mathbf{L}}{dt} + \vec{\omega} \times \mathbf{M}. \quad (2.9)$$

Reducing Equation (2.9) to fit the form of Equation (2.7), we find

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times (\mathbf{B}_0 - \vec{\omega}/\gamma). \quad (2.10)$$

Note that from the rotating frame perspective, as the frame precession approaches the Larmor frequency,  $\omega = \gamma B_0$ , the magnetization appears to be stationary.

The phenomenon of resonance occurs with the application of a transverse magnetic field,  $\mathbf{B}_1$ . This field must oscillate at a frequency  $\omega_0$  in order to tip the nuclei into a higher energy state. Such an oscillating field can be constructed from two circularly polarized fields rotating in opposite directions.

$$2B_1 \cos(\omega_0 t) = B_1 e^{-j\omega_0 t} + B_1 e^{+j\omega_0 t}$$

If  $B_1 \ll B_0$ , then only the component rotating in the same sense as the magnetization needs to be considered. This allows the transverse magnetic field to be written as

$$\mathbf{B}_1(t) = B_1 \cos(\omega_0 t) \hat{a}_x - B_1 \sin(\omega_0 t) \hat{a}_y$$

Rewriting Equation (2.7) with both the longitudinal,  $\mathbf{B}_0$ , and transverse,  $\mathbf{B}_1$ , magnetic fields one finds

$$\begin{aligned} \frac{dM_x}{dt} &= \gamma[M_y B_0 + M_z B_1 \sin(\omega_0 t)] \\ \frac{dM_y}{dt} &= \gamma[M_z B_1 \cos(\omega_0 t) - M_x B_0] \\ \frac{dM_z}{dt} &= \gamma[-M_x B_1 \sin(\omega_0 t) - M_y B_1 \cos(\omega_0 t)] \end{aligned}$$

with the solution

$$\begin{aligned} M_x &= M_0 \sin(\omega_1 t) \sin(\omega_0 t) \\ M_y &= M_0 \sin(\omega_1 t) \cos(\omega_0 t) \\ M_z &= M_0 \cos(\omega_1 t) \end{aligned}$$

where  $\omega_0 = \gamma B_0$  and  $\omega_1 = \gamma B_1$ . This stationary frame solution shows that the magnetization tends to precess about both fields at the rates  $\omega_0$  and  $\omega_1$  respectively. The effect of the transverse field is more easily seen by shifting one's perspective to the frame rotating about the z-axis at  $\omega_0$ . From this vantage point, the magnetization appears to precess only about  $\mathbf{B}_1$ . Thus, the effect of this transverse magnetic field is to rotate the net magnetization vector  $\mathbf{M}$  away from the z-axis. This rotation occurs in the plane orthogonal to the applied field  $\mathbf{B}_1$ . The angle of rotation  $\theta$  is controlled by the magnitude of the applied field and the length of time it is applied.

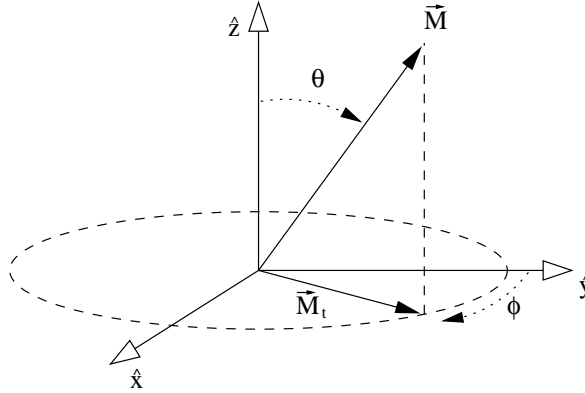


Figure 2.1: Magnetization vector and transverse plane projection

Typically, the physical coils that are used to apply the transverse field  $\mathbf{B}_1$  are the same coils used to acquire the imaging data. Thus, the magnitude of the received/recorded signal is proportional to the component of the magnetization that lies in the transverse plane. Applying a “90° pulse” rotates the magnetization vector completely into the transverse plane, and is typically the first step in the imaging process.

### 2.2.2 Relaxation

The previous section described the behavior of the bulk magnetization in the presence of magnetic fields. This section examines the dynamics once an applied transverse magnetic field has been removed. It takes a finite amount of time for the magnetization to return to the equilibrium state, parallel to the static magnetic field. The process of magnetization decay is called *relaxation*, and occurs primarily through two mechanisms.

The restoration of thermal equilibrium occurs primarily through a loss of energy between the spin system and the surrounding thermal reservoir - often termed the *lattice*. This process is known as *spin-lattice* or *longitudinal* relaxation. The mathematical description of the process is given by

$$\frac{dM_z}{dt} = -(M_z - M_0)/T_1 \quad (2.11)$$

with the solution

$$M_z(t) = M_z(0)e^{-t/T_1} + M_0(1 - e^{-t/T_1}).$$

The parameter  $T_1$  is often referred to as the spin-lattice relaxation time. The time it takes for the bulk magnetization to realign with the static field  $\mathbf{B}_0$  after excitation from a transverse pulse is dictated by  $T_1$ .

There is a secondary relaxation phenomena that occurs when the nuclear spins decay into thermal equilibrium among themselves. This occurs through state transitions between quantum states with similar energy. This process is known as *spin-spin*, or *transverse*, relaxation and is characterized by the time constant  $T_2$ . Interactions between particles with coupled-states affect the phase coherence of the aggregate collection of nuclear spin states. The strength of the magnetization vector depends on this coherence and a loss of coherence puts the magnetization out of focus. For completely incoherent spins, the net magnetization is zero. The time constant  $T_2$  is a measure of transverse magnetization loss due to the dephasing of the nuclear spins. The value of  $T_2$  for a given material is typically much less than  $T_1$ .

Analytically, the transverse relaxation process is given by

$$\frac{dM_{x,y}}{dt} = -M_{x,y}/T_2 \quad (2.12)$$

with the solution

$$M_{x,y}(t) = M_{x,y}(0)e^{-t/T_2}.$$

Dephasing through transverse relaxation can also be viewed from a classical perspective. Each magnetic moment in the population has a magnetic field that affects the neighboring moments as  $\sim \mu/r^3$ . Thus, a proportional difference in the magnetic field must be considered for each magnetic moment in the collection. This difference causes two moments to differ in precession by  $\delta\omega_0$  and after a time  $t = (\delta\omega_0)^{-1}$  the moments will be one radian out of phase. This loss of phase coherence causes the magnetization to lose focus and, subsequently, observed amplitude.

In the nuclear magnetic resonance literature, descriptions of the dynamics of an aggregate system of particles refer to two dephasing relaxation constants. One,  $T_2$ , refers to the non-recoverable energy lost from the system through dephasing. The other,  $T_2^*$ , is recoverable. The two types follow from the interpretation of the quantum physics description of spin density. Energy is recoverable if the spin distribution moves from one quantum state to another with equal energy. However, if there is no phase coherence between the two states, dephasing will occur although no energy is lost. Energy is only lost when the spin distribution moves to a lower energy state in the system distribution of spins and the dephasing in this case is non-recoverable.

### 2.2.3 The Bloch equations

Combining Equations (2.7), (2.11), and (2.12) in a rotating frame yields a system of equations known as the *Bloch equations*. Starting from Equation (2.7), and including terms describing relaxation effects, one can write

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B} - \frac{(M_x \hat{a}_x + M_y \hat{a}_y)}{T_2} - \frac{(M_x - M_z) \hat{a}_z}{T_1} \quad (2.13)$$

The magnetic field is comprised from both static and oscillating components. In the rotating frame of reference, this can be written as

$$\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_0 - \vec{\omega}/\gamma$$

This can be simplified by orienting  $\mathbf{B}_0$  along the z-axis, and rotating the frame at the rate  $\omega$ , such that the  $\mathbf{B}_1$  direction appears stationary along the x-direction. Under these assumptions, equation (2.13) can then written

$$\frac{d\mathbf{M}}{dt} = \gamma \begin{vmatrix} \hat{a}_x & \hat{a}_y & \hat{a}_z \\ M_x & M_y & M_z \\ B_1 & 0 & (B_0 - \omega/\gamma) \end{vmatrix} + \begin{bmatrix} M_x/T_2 \\ M_y/T_2 \\ (M_x - M_z)/T_1 \end{bmatrix}$$

Expanding the above component-wise one finds the Bloch Equations:

$$\frac{dM_x}{dt} = \gamma M_y (B_0 - \omega/\gamma) - \frac{M_x}{T_2}$$

$$\begin{aligned}\frac{dM_y}{dt} &= \gamma(-M_x(B_0 - \omega/\gamma) + M_z B_1) - \frac{M_y}{T_2} \\ \frac{dM_z}{dt} &= -\gamma M_y B_1 - \frac{(M_z - M_0)}{T_1}\end{aligned}$$

This model is useful for describing the dynamics of the MRI process to the first order. Note that as the rotating frame frequency,  $\omega$ , approaches the Larmor frequency, the  $\mathbf{B}_0$  terms disappear. With application of the  $\mathbf{B}_1$  field, the equations describe a rotation of the magnetization through the plane defined by the  $\hat{a}_z$  and  $\hat{a}_y$  directions.

#### 2.2.4 An example of spin manipulation: simple (Hahn) spin echo

The previous sections provide a basic analytical foundation for the behavior of material in a strong magnetic field. The additional application of radio-frequency pulses and magnetic field gradients can be used to manipulate the bulk magnetization to great effect, ultimately allowing the construction of images via non-invasive means. This section illustrates an example of such spin manipulation: the simple spin echo.

A simple spin echo is commonly used to overcome magnetic field inhomogeneity. Field inhomogeneity in the static field causes the magnetization to lose phase coherence over time. For a change  $\Delta B_0$  in the static field, the phase coherence time is inversely proportional to  $\gamma \Delta B_0$ . The central idea of the spin echo technique is to apply a  $180^\circ$  rf pulse that conjugates the orientation of the magnetization vector in the transverse plane at some time  $\tau$  after the initial  $90^\circ$  rf pulse that rotated the magnetization vector into the transverse plane. The effect of this second pulse is to place the magnetization vector ahead of the focusing point so that as the dephasing evolves, the magnetization vector refocuses again at time  $2\tau$ .

The entire pulse sequence can be succinctly described via the following diagram:

$$\begin{aligned}I_z &\xrightarrow{-\frac{\pi}{2}\hat{a}_x} I_y \xrightarrow{-\Delta\omega_0\tau\hat{a}_z} (I_y \cos \phi + I_x \sin \phi) \xrightarrow{-(\pi)\hat{a}_y} (I_y \cos \phi - I_x \sin \phi) \xrightarrow{-\Delta\omega_0\tau\hat{a}_z} \\ &\overline{[I_y \cos^2 \phi + I_x \cos \phi \sin \phi - I_x \cos \phi \sin \phi + I_y \sin^2 \phi]} = I_y\end{aligned}$$

where the arrows designate transitions in the spin state for the given operator, and  $\phi$  is the



precessional phase shift,  $\Delta\omega\tau$ . A description of this pulse sequence is given in Table 2.1.

$I_z$	At $t = 0$ , the magnetization vector is aligned with the static field $\mathbf{B}_0$ in the $\hat{a}_z$ direction.
$\xrightarrow{-\frac{\pi}{2}\hat{a}_x} I_y$	After application of a $90^\circ$ pulse in the minus $\hat{a}_x$ direction the magnetization vector is oriented in the $\hat{a}_y$ direction.
$\xrightarrow{-\Delta\omega_0\tau\hat{a}_z} (I_y \cos \phi + I_x \sin \phi)$	Due to inhomogeneities in the static field, the spins begins to lose phase coherence. The magnetization picks up both $\hat{a}_x$ and $\hat{a}_y$ components.
$\xrightarrow{-(\pi)\hat{a}_y} (I_y \cos \phi - I_x \sin \phi)$	After a time $\tau$ , a $180^\circ$ pulse in the minus $\hat{a}_y$ direction is applied. This has the effect of changing the polarity of the $\hat{a}_x$ components. It has no effect on the $\hat{a}_y$ components.
$\xrightarrow{-\Delta\omega_0\tau\hat{a}_z} I_y$	The magnetic field inhomogeneity continues to dephase the spin, at a rate $\Delta\omega t$ . However, the spins have been placed ahead of the refocusing point, so that a time $\tau$ from the $180^\circ$ pulse, the magnetization lands in back into coherence.

Table 2.1: Simple spin echo sequence

The simple spin echo described above is used in a number of imaging protocols for the express purpose of compensating for magnetic field inhomogeneity. It was presented here to provide a short example of how the bulk magnetization may be manipulated through the application of rf pulses.

### 2.2.5 Classical dynamics summary

This section sought to show that while the quantum mechanical behavior of matter is never far below the surface, the MR imaging process can be accurately described using dynamic models from classical physics. MRI builds upon the inherent physical property that magnetic moments of a material will precess when placed in a magnetic field. This section provided a description of this precession from a classical physics perspective. Furthermore, the precession phenomena allows the aggregate collection of magnetic moments to be manipulated in space, and allows for the possibility of overcoming the effects of relaxation and decoherence that are present in the imaging system. All of the imaging techniques that follow rely on magnetization manipulation to some degree.

## 2.3 Acquisition of an image

The previous section gave a review of both the modern and classical physics perspectives on the nature of atomic particles in a magnetic field. This section describes how the manipulation of such particles can generate an image. First, a description of the electro-magnetic signal measured by the imaging system is given. From this signal an image showing the location and composition of the particles can be reconstructed.

### 2.3.1 Signal detection

We first describe the data collection process. For reference, we assume that the static field,  $\mathbf{B}_0$ , is oriented along the positive  $z$  axis. At equilibrium, the net magnetic moment is parallel to this magnetic field.

If a coil is placed with its symmetry axis transverse to the static field  $\mathbf{B}_0$ , the precessing magnetization will induce an oscillating electromotive force (e.m.f.) at the Larmor frequency  $\omega_0$ . Only that component of the magnetization that lies in the transverse plane will induce current in the coil, so as the magnetization relaxes, the e.m.f. signal will decay. This is known as the *free induction decay* (FID). Through the Fourier transform, this signal can be represented in the frequency domain as very narrow band signal.

The decaying magnetization can easily be represented in complex number notation as

$$M_+(t) = M_0 e^{j\omega_0 t} e^{-t/T_2} \quad (2.14)$$

The e.m.f. signal detected in the coil is proportional to  $M_+$ . The received signal can be demodulated to a lower frequency band by mixing the received signal with a reference signal oscillating at  $\omega_r$ . The result of this *heterodyne* process is

$$S(t) = S_0 e^{-t/T_2} e^{j\Delta\omega t}$$

where  $\Delta\omega = \omega_0 - \omega_r$ .

### 2.3.2 Gradient fields, spin density, and $k$ -space

To discern FID signals from similar media at different locations, a magnetic field gradient is introduced. For example, if the central field of the experiment is a combination of a static field  $\mathbf{B}_0 = B_0 \hat{a}_z$ , and gradient field oriented along the  $z$ -axis  $\mathbf{G} = Gr \hat{a}_z$ , Equation (2.7) then becomes

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times (\mathbf{B}_0 + \mathbf{G}) = \gamma \mathbf{M} \times [(B_0 + Gr) \hat{a}_z].$$

Given that the magnetic field varies linearly along  $r$ , the Larmor Frequency varies with  $r$  as well,

$$\omega(r) = \gamma B_0 + \gamma Gr.$$

This simple linear relation between the Larmor frequency and the nuclear spin coordinates lies at the heart of magnetic resonance imaging. Along  $r$ , similar media precess at slightly varied frequencies due to the gradient field  $\mathbf{G}$ . The value of  $\Delta\omega$  in the received e.m.f. signal is then used to map the inductive magnetization to a location on the  $r$  axis.

In general, the magnetization can be described as the summation of the spin density over a small volume,  $\rho(\mathbf{r})dv$ . From the equation describing free induction decay (2.14), and recognizing that the signal received is proportional to the transverse magnetization, the signal received from the spin density region is

$$dS(t) = (S_0 e^{t/T_2}) \rho(\mathbf{r}) dV e^{j(\gamma B_0 + \gamma \mathbf{G} \cdot \mathbf{r})t}$$

Neglecting for the moment the relaxation decay and demodulating at a frequency  $\omega_0$  to remove the static field contribution from the expression we find

$$dS(t) = \rho(\mathbf{r}) dV e^{j\gamma \mathbf{G} \cdot \mathbf{r}t}.$$

After integrating, we find that the spin density  $\rho(\mathbf{r})$  and the received signal are related as

$$S(t) = \int_V \rho(\mathbf{r}) e^{j\gamma \mathbf{G} \cdot \mathbf{r}t} d\mathbf{r} \quad (2.15)$$

By defining a reciprocal spatial term  $\mathbf{k}$  as

$$\mathbf{k} = \frac{1}{2\pi} \gamma \mathbf{G}t, \quad (2.16)$$

Equation (2.15) is recognized as the Fourier Transform

$$S(\mathbf{k}) = \iiint \rho(\mathbf{r}) e^{j2\pi\mathbf{k}\cdot\mathbf{r}} d\mathbf{r} \quad (2.17)$$

$$\rho(\mathbf{r}) = \iiint S(\mathbf{k}) e^{-j2\pi\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} \quad (2.18)$$

Thus, the spin density of the material and the received free inductive decay signal are mutually related. Sampling occurs along each dimension of  $\mathbf{k}$  space. Performing an inverse Fourier transform on this sampled  $k$ -space data gives a description of the spatial composition of the space scanned by  $\mathbf{r}$ . From this data, images can be constructed.

### 2.3.3 Selective excitation

The applied transverse rf pulse,  $\mathbf{B}_1$ , oscillating at  $\omega$ , affects only a specific region in the sample due to the resonance phenomenon. The applied field can isolate either a chemical composition or a spatial slice through the material. Spatial resolution is restricted by a time-frequency relationship, with the bandwidth BW of the signal inversely proportional to the pulse duration T, i.e.,

$$\Delta BW \propto \frac{1}{\Delta T}$$

Functionally, there are two classes of pulses, *hard* and *soft*. Hard pulses are intense broadband excitations, typically of very short time and consequently broad in bandwidth. Soft pulses are weak, narrow-band signals. Three general types of modulated pulses are

the Rectangular Pulse   Produces a sinc excitation profile in the frequency domain.

the Gaussian Pulse   Provides a smooth envelope between off and on states. Gaussian pulses are typically used to remove the side lobes in a frequency excitation profile.

the Sinc Pulse   Produces a rectangular profile in the frequency domain with some ringing.

Selective excitation can also be achieved through a combination of hard pulses. For a given hard pulse of width  $\Delta t$  driven at  $\omega_0$ , those particles with a resonant frequency  $\omega_0$  will tip farther into the

transverse plane than those particles that are off-frequency. By using  $m$  successive series of pulses, separated in time by  $\tau$ , the resonant particles will be forced to a tip angle  $\theta$  while the non-resonant particles remain relatively unchanged. This idea was presented by Morris and Freeman, [30], and is named the DANTE sequence which “alludes the repetitive circular journeys by Dante and Virgil in Dante Alighieri’s *Purgatorio*, akin to the trajectories undergone by off-resonant spins.” [5].

$$\text{Total Flip Angle} = \theta = m(\gamma B_1)\Delta t$$

This ability to approximate soft pulse profiles through a series of applied hard pulses is fundamental to the linear system model described in Section 3.2. This system model provides the foundation for the low-order imaging methods presented later in this thesis.

### 2.3.4 2-D Fourier imaging

Traditional Fourier imaging uses the manipulation of gradient fields and the application of rf pulses to extract a signal from a given slice within a tissue volume. Typically, an rf pulse is used to select the slice. For each acquisition, one gradient field is used to scan one line of  $k$ -space. This gradient is typically referred to as the *read* gradient. A second gradient field is used to position the line to read. It does this by synchronizing the phase along the axis orthogonal to the read gradient. Thus, this second field is referred to as the *phase* gradient.

Setting the read gradient as  $G_x$  and the phase gradient as  $G_y$ , Equation (2.18) can be represented as

$$S(k_x, k_y) = \int_{-a/2}^{a/2} \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y, z) e^{j2\pi(k_x x + k_y y)} dx dy \right] dz \quad (2.19)$$

where  $a$  is the slice thickness. The integral over the  $dz$  region represents an averaging process over the whole slice. The term in brackets is the two dimensional Fourier transform of the spin area density. A timing diagram showing the relative placement of the gradients and rf pulses in time are given in Figure 2.2. As shown in the timing diagram, the  $90^\circ_x$  soft pulse is used to tip the magnetization vector into the transverse plane. A  $180^\circ_y$  pulse is then used to refocus the

magnetization.

The parameters  $k_x = \frac{1}{2\pi}\gamma G_x t_x$  and  $k_y = \frac{1}{2\pi}\gamma G_y t_y$  refer to different periods in the acquisition sequence and relate to different gradients. First the location along  $k_y$  is selected by setting the phase gradient  $G_y \neq 0$  and  $G_x = 0$ . In this case the spins evolve along the positive y-axis in  $k$ -space. The location along  $k_y$  can be set by either a fixed gradient applied over a variable length of time or by using an adjustable gradient magnitude for a fixed length of time. To begin the signal acquisition, first the phase gradient is switched off,  $G_y = 0$ , and the read gradient is switched on,  $G_x \neq 0$ . In this case the spins evolve along the positive x-axis in  $k$ -space, and the data sampling occurs at the y-axis intercept set by  $t_y$  or  $G_y$ . Figure 2.3 shows this graphically.

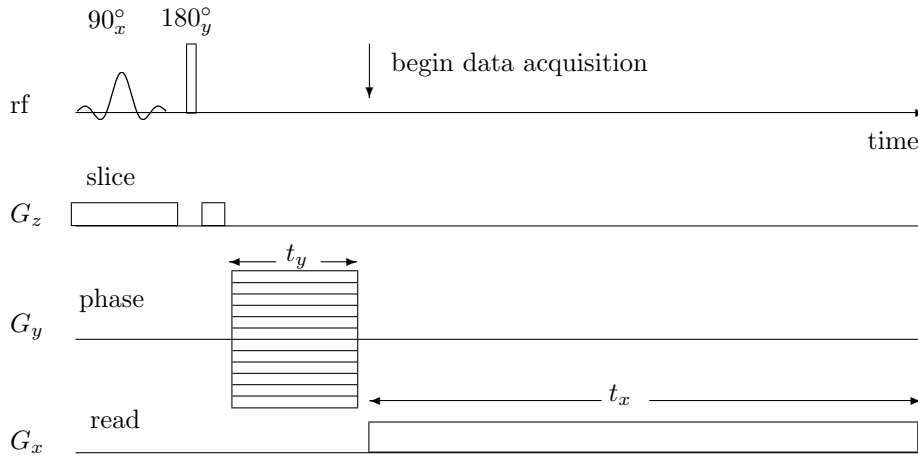


Figure 2.2: 2-D Fourier Acquisition Timing Diagram (from [5])

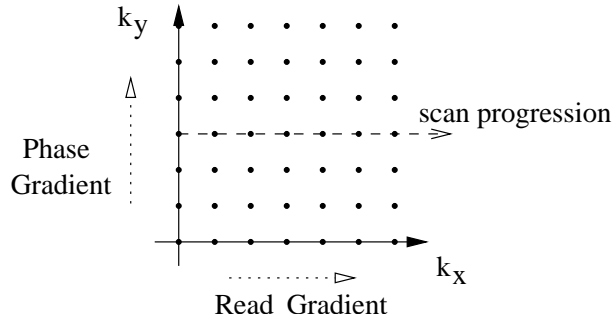


Figure 2.3: Using magnetic field gradients to scan  $k$ -space

## 2.4 Summary

This section provided a brief description of the fundamental physical models used to understand the magnetic resonance imaging process. The ability to image tissue non-invasively using MRI begins with the concept of *spin*, an intrinsic property of all matter. While the quantum nature of spin is the fundamental mechanism of imaging, an aggregate collection of spins can be modeled using classical dynamics and manipulated through the application of rf frequency electro-magnetic pulses and magnetic field gradients. Using these forms of interaction, the tissue structure and composition can be encoded. After the excitation pulses are removed, the system of spins induces a signal in a coil transverse to the static magnetic field as it relaxes. This signal is sampled and produces a  $k$ -space description of the encoding. From this sampled data images of the tissue can be formed. This section closed with a description of the traditional 2-D Fourier imaging method. The next section details methods to acquire images with a minimum amount of sampled data using both Fourier and non-Fourier based imaging techniques.

## Chapter 3

# Image acquisition via low order encoding

The basic pretext for low-order imaging is that in dynamic MRI sequences, only a small part of the image changes from frame to frame. The goal then is to acquire a limited amount of data at each image sampling instant, and reconstruct an *estimate* of the image guided by some prior knowledge of the image sequence. Typically, this includes using some combination of the most recently acquired data with data from a reference image to construct the image estimate. The advantage of low-order encoding is that for many image acquisition protocols the image acquisition time is proportional to the number of sampled  $k$ -space lines. Thus, if one can reduce the number of lines required to reconstruct an image, one can reduce the image acquisition time.

This chapter presents a review of three methods that are the most successful application of this simple idea to date. Fourier Keyhole, the subject of Section 3.1.1, was proposed first and is the most straight forward of the three methods. Reduced encoding methods such as RIGR and singular value decomposition (SVD) techniques, the topics of § 3.1.2 and § 3.2.1 respectively, soon followed. The new methods presented in Chapters 4 and 5 build upon the linear system model that



is central to the SVD method. Thus, the section closes with a detailed discussion of this model.

### 3.1 Fourier based methods

The following low-order acquisition methods are derived from traditional Fourier imaging techniques.

#### 3.1.1 Fourier Keyhole

The Fourier Keyhole (FK) method was proposed by Brummer and Van Vaals, et. al. [4, 43], and results from the following simple concept. In MRI images, a significant percentage of the signal energy is contained in the lower frequency components of  $k$ -space. Thus, one can expect a reasonable estimate of the image if one acquires a limited number of low-frequency  $k$ -space lines and fills out the  $k$ -space data matrix using data from a reference image.

Analytically, this can be described as follows. Using rf input signals, a slice located at  $z = z_0$  is selected and through the manipulation of the magnetic field gradients the received signal at time  $t$

$$S(k_x, k_y, t)|_{z=z_0} = \int_{-\infty}^{\infty} \rho(x, y, t) e^{j2\pi(k_x x + k_y y)} dx dy. \quad (3.1)$$

is sampled for a range of  $k_x$  and  $k_y$  values to construct the  $k$ -space data matrix. For the reference image, this sampling is performed over the ranges  $-N/2 < k_y < N/2$  and  $-M/2 < k_x \leq M/2$ . Subsequent images are then acquired by sampling only a limited range of  $k$ -space along one direction, for example  $-r/2 < k_y \leq r/2$  with  $r < M$ , and replacing that range in the sampled  $k$ -space matrix of the reference data.

$$S(k_x, k_y, t)|_{z=z_0} = \begin{cases} \int_{-\infty}^{\infty} \rho(x, y, t) e^{j2\pi(k_x x + k_y y)} dx dy, & \text{for } |k_y| \leq r/2 \\ S(k_x, k_y, 0), & \text{for } |k_y| > r/2 \end{cases}$$

The sampled  $k$ -space data can be represented in discretized form as a data matrix  $\mathcal{R}_t$ . From a linear algebra perspective, acquiring the lowest frequency components of the system response is equivalent to selecting columns from the  $k$ -space data matrix that are associated with the lowest

frequency components of the Fourier basis, i.e.,  $\mathcal{R}_t I_{n,p}$  where  $I_{n,r}$  are  $r$  columns from the identify matrix of size  $n$ . This allows a linear algebra version of the algorithm to be described as

$$\hat{\mathcal{R}}_t = \mathcal{R}_t I_{n,p} I_{n,p}^T + \mathcal{R}_0 (I_n - I_{n,p} I_{n,p}^T)$$

The matrix algebra description of the FK image estimation algorithm is given in Table 3.1.

**Fourier Keyhole Dynamic Sequence Acquisition Method**

```

 $\mathcal{R}_0 = k$ -space data matrix of reference image
for each new acquisition
     $\mathcal{R}_t = k$ -space data matrix of image at time  $t$ 
     $\hat{\mathcal{R}}_t = \mathcal{R}_t I_{n,p} I_{n,p}^T + \mathcal{R}_0 (I_n - I_{n,p} I_{n,p}^T)$ 
end

```

Table 3.1: Fourier keyhole dynamic sequence acquisition method

The FK method has been shown to be quite effective in estimating contrast change sequences [43]. The effectiveness of the Fourier keyhole method will be analyzed in more detail in Section 5.3.

### 3.1.2 Reduced-encoding imaging via generalized-series reconstruction (RIGR)

The Reduced-encoding Imaging via Generalized-series Reconstruction (RIGR) method was proposed in 1994 by Liang and Lauterbur [25] and is an extension of the Fourier keyhole method described in the previous section. The central concept of the method is to identify a linear combination of the central region  $k$ -space basis functions that most accurately reflect the phase-encoded data in the central region of  $k$ -space. The model parameters identified in this first step are then used to estimate the unmeasured phase-encoded data to fill-out the rest of the  $k$ -space data matrix.

For  $r$  lines of sampled central region  $k$ -space data, the estimate may be written as

$$\hat{\rho}_{dyn}(u, v) = |\rho_{ref}(u, v)| \circ \sum_{n=-r/2}^{r/2-1} c_n e^{j2\pi n \Delta k u} \quad (3.2)$$

where  $u$  and  $v$  are the indices of the sampled spin density matrix,  $c_n$  are the RIGR model parameters, and  $\circ$  is an element-by-element product (also known as the Hadamard product or Schur product, [18, Chp. 5]). This estimation step is performed on a row-by-row basis to construct the estimate of the dynamic image. The model parameters are determined via

$$d_{dyn}(m, v) = \sum_{n=-r/2}^{r/2-1} c_n \hat{d}_{ref}(m-n, v) \quad -r/2 \leq m \leq r/2-1 \quad (3.3)$$

where

$$\hat{d}_{ref}(m-n, v) = \int_{-\infty}^{\infty} |\rho_{ref}(u, v)| e^{-j2\pi(m-n)\Delta k u} du. \quad (3.4)$$

This set of equations identifies the model parameters  $c_n$  via a best linear fit of the reference data to the most recently sampled data.

Note that the estimated image data in (3.2) results from a Schur product of the reference image with a linear combination of the central-region  $k$ -space basis functions. In effect, this imposes a spatial envelope profile over the estimated data points and is the true strength of the method. As shown in the examples of Chapter 5 and [13], the RIGR method is very effective in imaging contrast change sequences. However, it is limited by a bias towards the spatial composition of the reference image, and is quite unsuitable for sequences exhibiting motion change or image sequences displaying high intensity pixels in regions that were very low intensity in the reference image. The effectiveness of the RIGR method will be explored in more detail in Section 5.3.

Table 3.2 gives a description of RIGR from a matrix algebra perspective.

### 3.2 A linear system model for non-Fourier based methods

Traditional Fourier imaging uses successive rf pulses to select slices, and then uses gradient manipulation of the spins to sample the two dimensional  $k$ -space signal from the sample, as described in Section 2.3.4. This section describes a different technique to acquire the same  $k$ -space data. Specifically, one may use non-Fourier encoding techniques to sample a plane in  $k$ -space at a fixed point  $k_{z0}$ . The material that follows was drawn primarily from [36].

**Reduced Encoding by Generalized Series Reconstruction (RIGR) Method**

Let  $I_{n,r}$  be the  $r$  columns of the identity matrix that capture the lowest frequency components of the  $k$ -space data matrix  $\mathcal{R}_t$  of size  $m \times n$ . Let  $\mathcal{X}$  be the sampled versions of those same low frequency components of the Fourier basis set.

```

 $\mathcal{R}_0$  =  $k$ -space data matrix of reference image
 $r$  = number of  $k$ -space data lines to acquire
 $ctr = m/2 + 1$ , a count holder for the  $k$ -space data matrix corresponding to  $\omega = 0$ 
 $d_0 = \mathcal{R}_0 I_{n,r}$ 
for each new acquisition
     $\mathcal{R}_t$  =  $k$ -space data matrix of image at time  $t$ 
     $d = \mathcal{R}_t I_{n,p}$ 
    for each column  $v$  in  $\hat{\mathcal{R}}_t$ 
         $H = \text{toeplitz}(d_0(ctr : ctr + r - 1, v), d_0(ctr : -1 : ctr - r + 1, v))$ 
         $c = H^{-1}d(:, v)$ 
         $\hat{\mathcal{R}}_t(:, v) = \mathcal{R}_0(:, v) \circ (\mathcal{X}c)$ 
    end
end

```

Table 3.2: RIGR dynamic sequence acquisition method

As shown in Section 2.3.3, soft or hard pulses can be used to excite the magnetization. In practice, soft pulses can be approximated by piece-wise-linear hard pulses. In the limit that these hard pulses become infinitely narrow, but separated by a time  $\Delta t_p$ , they can still be used to excite the magnetization in the same manner as a continuous soft pulse. This sequence of hard pulses can be described by

$$p_H(t) = \sum_n p_n \delta(t - n\Delta t_p)$$

where the individual pulses can be complex valued. The phase component of the pulses relates to the relative position of the magnetization at the onset of the rf pulse. Note as well that the following relationship holds: A narrow pulse in time gives a broad band signal in the temporal Fourier space; this in turn translates to a wide excitation profile in the spatial domain; which in turns translates to a narrow band in the spatial Fourier, or  $k$ -space, domain.

In the theoretical limit, such pulses can be represented by the Dirac delta function. Such pulses impart energy that flips the spins “instantly” at time  $t$ , after which the spins undergo free

precession in the time interval  $\Delta t_p$ . The total signal from all spins at time  $\tau$  due to the  $n^{\text{th}}$  hard pulse is

$$S(k_x, k_y, k_n) = \iiint \rho(x, y, z) (p_n e^{-j k_n z}) e^{-j(k_x x + k_y y)} dx dy dz. \quad (3.5)$$

The spatial encoding in  $k$ -space is related to the gradients by

$$\begin{aligned} k_n &= \gamma G_z n \Delta t_p && \text{phase encoded in } k_z \\ k_y &= \gamma G_y T && \text{phase encoded in } k_y \\ k_x &= \gamma G_x \tau && \text{signal read along } k_x \end{aligned}$$

where  $T$  is the duration of the phase encoding gradient pulse. Note that this formula is valid for any tip angle, as long as the axial length of the sample is shorter than the spatial period in  $z$ , or equivalently,

$$\text{sample length in } z < \frac{1}{\gamma G_z \Delta t_p}.$$

For small flip angles,  $\sin \theta \approx \theta$  and the Bloch equations can be accurately approximated to the first order. One can then apply superposition to remove the dependence on  $n$  in the received signal.

$$\sum_n S(k_x, k_y, k_n) = S(k_x, k_y) = \iiint \rho(x, y, z) \left[ \sum_n p_n e^{-j k_n z} \right] e^{-j(k_x x + k_y y)} dx dy dz. \quad (3.6)$$

The quantity in brackets is the magnetization profile and is equal to the Fourier transform of the excitation series  $\{p_n\}$ . Note that in this equation, off-resonance and  $T_2$  relaxation effects are ignored. The superposition mechanism is thus only valid if the evolution due to these effects occurs in a time much less than the time between rf pulses.

Superposition can also be used to build a system response model. If using only low-flip angles, the received signal from a given pulse can be constructed from a superposition of known hard pulse responses. The excitation rf pulse can be computed as a linear combination of pulses. Thus it should be possible to construct the response of a system to an input  $p(t)$

$$p(t) = \sum_m g_m c_m(t) \quad m = 1 \dots M$$

if the responses to the input set  $\{c_m(t)\}$  are known. Using the set of responses and the weighting coefficients  $g_m$ , one can construct the following matrices

$$C = \left[ \begin{array}{c|c|c|c} c_1(t) & c_2(t) & \cdots & c_M(t) \end{array} \right] \quad g = \begin{bmatrix} g_1 \\ \vdots \\ g_M \end{bmatrix}.$$

The input pulses  $c_m$  can be represented by digital samples rather than continuous functions by the following transformation.

$$c_m(t) = \sum_n c_{m,n} \Pi_n(t - n\Delta t_p)$$

where  $\Pi$  is the rf unit-box pulse.

$$\Pi_n(t) = \begin{cases} \text{constant,} & n\Delta t_p < t < (n+1)\Delta t_p \\ 0, & \text{otherwise} \end{cases}$$

The accumulated rf pulse response can then be written as a sum of these unit box functions.

$$p(t) = \sum_m \sum_n g_m c_{m,n} \Pi_n(t)$$

or in discrete form

$$p_m = \sum_n g_n c_{m,n} \iff \mathcal{P} = Cg$$

Generally, any received signal sampled in time can be represented as a discrete sequence  $\{y_k\}$ . Let  $\mathcal{R}_n(t)$  or  $\mathcal{R}_{n,k}$  be the response from the box-pulse excitation function  $\Pi_n(t)$ . Then the mapping between the input and response of the system is described by  $\mathcal{R}$

$$p(t) \xrightarrow{\mathcal{R}(t)} y(t) \iff p_k \xrightarrow{\mathcal{R}_k} y_k$$

This mapping can be described with a matrix notation as follows

$$y_k = \sum_n p_n \mathcal{R}_{n,k} = \sum_n \sum_m g_m c_{m,n} \mathcal{R}_{n,k}$$

or

$$\mathcal{Y} = \mathcal{R}Cg = \mathcal{R}\mathcal{P}.$$

Note that in this context,  $\Pi_n$  acts as a delta function, and  $\mathcal{R}$  is the system impulse response matrix. Also,  $\mathcal{R}$  is not shift invariant, otherwise a single  $\Pi$  could be used to describe it.

From this matrix representation, a tissue sample can be imaged through non-Fourier techniques. The ability to rotate the collection of input vectors to a new basis set, unrelated to the Fourier basis that dominates traditional imaging, opens up a wide range of imaging modalities. The received signal recorded during an imaging experiment will contain data from the tissue sample that is supported by the sub-space spanned by the input basis. This allows wavelet or SVD based techniques to be used in multiple rf scan experiments [33, 36].

### 3.2.1 SVD encoding method

The SVD method proposed by Panych and Zientara, et. al. [47, 34], is conceptually very simple. To acquire a dynamic sequence, one uses rf encoding and a low magnetization tip angle which allows one to model the image acquisition process using the linear system model described above. The full  $k$ -space data matrix of the first image is acquired. The SVD of this data matrix is calculated (A.2), and the dominant singular vectors are used to acquire and reconstruct the subsequent images in the sequence. If the matrix  $\mathcal{P}$  is composed of columns from the right singular vectors of  $\mathcal{R}$  [47], then an estimate of the system response matrix can be constructed via

$$\hat{\mathcal{R}} = \mathcal{Y}\mathcal{P}^H = \mathcal{R}\mathcal{P}\mathcal{P}^H. \quad (3.7)$$

The SVD image estimation algorithm is given in Table 3.3. Variants of the SVD method are given in Section 5.1.

### 3.2.2 The relationship between spatial and $k$ -space representations of an image

As shown in Section 3.2.1, the MRI imaging process can be described by a linear system under certain conditions [47]. Specifically, spatial encoding by manipulation of spatially selective radio-

**Singular Value Decomposition (SVD) Dynamic Sequence Acquisition Method**

$\mathcal{R}_0$ , the  $k$ -space data matrix of reference image  
 $\mathcal{R}_0 = \mathcal{U}\Sigma\mathcal{V}^H$ , singular value decomposition  
 $\mathcal{P} = \mathcal{V}(:, 1 : r)$ , the input vectors for the sequence  
**for each new acquisition**  
 $\mathcal{R}_t = k$ -space data matrix of image at time  $t$   
 $\hat{\mathcal{R}}_t = \mathcal{R}_t\mathcal{P}\mathcal{P}^H$   
**end**

Table 3.3: SVD dynamic sequence acquisition method

frequency (rf) profiles together with small-flip-angle excitations allow one to analytically describe the imaging process as a linear system [36]. Thus, if an input rf-encoding excitation sequence is described by  $\mathcal{P}$ , then the output  $\mathcal{Y}$  of the imaging experiment can be described by

$$\mathcal{Y} = \mathcal{R}\mathcal{P}$$

where  $\mathcal{R}$  is an  $N \times N$  system matrix representation of the soft tissue response.

The linear system response model description developed by Panych, et. al., [36] spoke primarily towards sampling  $k$ -space directly. The focus of our research is the acquisition and tracking of data in the spatial (or image) domain. Mapping data between the two domains is easily accomplished by defining the  $N \times N$  unitary Fourier transform matrix [19, Chp. 5]

$$F_N = \left\{ (N)^{-1/2} e^{-j2\pi kn/N} \right\}, \quad 0 \leq k, n \leq N-1. \quad (3.8)$$

This allows one to transform the sampled  $k$ -space data matrix  $\mathcal{R}$  to the image matrix  $A$  via

$$A = F_M^H \mathcal{R} F_N. \quad (3.9)$$

The  $k$ -space sampling and output vectors can be transformed to the spatial domain in a similar way, via  $X = F_N^H \mathcal{P}$  and  $Y = F_N^H \mathcal{Y}$ . For the problems presented below, we choose to work entirely in the spatial image domain. The linear model used throughout the remainder of the thesis is

$$Y = AX \quad (3.10)$$



where  $X$  and  $Y$  may describe a single rf-encode excitation, i.e.,  $X$  and  $Y$  are column vectors, or a collection of multiple excitation experiments, i.e.,  $X$  and  $Y$  are matrices whose columns are input or output vectors respectively.

Note that the matrix transform given in (3.9) is *not* the traditional two-dimensional Discrete Fourier Transform (2D-DFT), which is defined as  $A = F_M \mathcal{R} F_N$ . The only significant effect of choosing  $F_M^H$  rather than  $F_M$  for the left matrix operator is to reverse the order of the basis vectors, in a sense running the frequency basis index  $k$  in the positive (+) direction rather than the negative (−) direction. Although the transformation is similar, (3.9) was chosen because it provides a frequency domain to spatial domain transform that is consistent for both left and right vector multiplication. For example, the singular value decomposition of  $\mathcal{R}$  is defined as

$$\mathcal{R} = \mathcal{U} \Sigma \mathcal{V}^H,$$

where  $\mathcal{U}, \mathcal{V}$  are unitary matrices and  $\Sigma$  is a diagonal matrix containing the singular values,  $\sigma_i$ , ordered in decreasing order. Transforming  $\mathcal{R}$  to the spatial domain via (3.9), one finds

$$\begin{aligned} A = F_M^H \mathcal{R} F_N &= F_M^H \mathcal{U} \Sigma \mathcal{V}^H F_N \\ &= (F_M^H \mathcal{U}) \Sigma (F_N^H \mathcal{V})^H \\ A &= U \Sigma V^H \end{aligned}$$

which gives the SVD of the spatial domain data as expected.

### 3.3 Useful error measures

For low-order imaging methods, such as those listed previously in this section, the decrease in dynamic MRI sequence acquisition time is a result of estimating the image rather than acquiring the full image data set. To measure the quality of the image estimates, we use the following error criteria.

### 3.3.1 Measuring distance between images and estimates

If  $\hat{A}$  is a given estimate of the true image  $A$ , then one typically would measure the *error* between the two using the Frobenius norm of the difference matrix [18],

$$\mathcal{E} = \|A - \hat{A}\|_F^2 = \sum_i \sum_j (a_{ij} - \hat{a}_{ij})^2, \quad (3.11)$$

where  $a_{ij}$  and  $\hat{a}_{ij}$  are the matrix elements at the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $A$  and  $\hat{A}$ , respectively.

An extension of this error measure is to determine the *relative error* of the estimate via  $\text{re}(\hat{A}, A) = \|A - \hat{A}\|_F^2 / \|A\|_F^2$ . For the region of interest (ROI) acquisition problems discussed in Chapter 4, we define a selection matrix  $S$  with elements  $s_{ij} = \{0, 1\}$ . The ROI is identified by the non-zero region of the selection matrix. The relative error measure thus becomes

$$\text{re}(\hat{A}, A, S) = \frac{\|S \circ (A - \hat{A})\|_F^2}{\|S \circ A\|_F^2}, \quad (3.12)$$

where  $\circ$  describes an element-by-element matrix product.

### 3.3.2 Measuring distance between subspaces

In the dynamic problems discussed in Chapter 5, the main concern is the ability to identify the various subspaces of the underlying image. Thus, we calculate the *principal angles* between dominant subspaces as a second criterion to evaluate the quality of image estimates in a dynamic sequence. The principal angles,  $\theta_k \in [0, \pi/2]$ , between two subspaces  $C$  and  $D$  are recursively defined [2] for  $k = 1, 2, \dots, r$  by

$$\cos \theta_k = \max_{u \in C} \max_{v \in D} u^H v = u_k^H v_k, \quad \|u\|_2 = 1, \|v\|_2 = 1,$$

subject to the constraints

$$u_j^H u_k = 0, \quad v_j^H v_k = 0, \quad j = 1, 2, \dots, k-1.$$

The vectors  $u_j$  and  $v_j$  need not be uniquely defined, but the principal angles always are.

There are a variety of methods to calculate principal (or canonical) angles [2, 40]. The most convenient method is to compute the singular value decomposition of the cross-correlation matrix

of the subspaces. For example, consider two orthonormal tall-and-thin matrices  $V_C$  and  $V_D$  of size  $N \times r$  with  $r < N$ . Each describes a subspace in the larger Euclidean space of all  $N \times N$  matrices. The principal angles between the two subspaces can be found through the SVD of  $M = V_C^H V_D = U_M \Sigma_M V_M^H$ . Specifically, the principal angles are  $\theta_i = \cos^{-1}(\sigma_M)_i$ . It should be noted that this method is fast, but not very accurate for angles close to zero, or equivalently, for singular values of  $M$  that are close to one.

### 3.4 Summary

This section described in some detail the fundamental principles behind low-order acquisition of dynamic MRI sequences. A review of the Fourier Keyhole (FK), Reduced Encoding via Generalized-Series Reconstruction (RIGR), and SVD methods was provided. In addition, this section provided a complete development of the linear system model that is fundamental to the SVD method. This linear system model forms the foundation of each of the imaging methods described in the remainder of this thesis.

## Chapter 4

# Efficient region of interest acquisition

As mentioned previously, for most dynamic MRI sequences the medically significant changes that occur between frames are often localized to a small region of interest (ROI). Thus, this section examines the efficient reconstruction of a pre-specified and arbitrarily shaped ROI. The problem examined below seeks to identify the most efficient set of data acquisition and image reconstruction vectors for a given static image and ROI. It is presumed that solutions to this static problem will be useful in guiding solutions to dynamic ROI acquisition problems.

### 4.1 Problem formulation

From the foundation of the linear system response model given in Section 3.2.1 above, the problem approached in this section is to acquire and represent only certain elements of the true image matrix  $A$ . In particular, we adopt the outer-product machinery,  $XL^H$ , suggested by the SVD method described in Section 3.2.1, but choose  $X$  and  $L$  to reconstruct a specified but arbitrarily shaped region of interest within the image matrix. The elements of interest are described through

an  $M \times N$  *selection matrix* matrix  $S$ , with elements  $s_{ij} \in \{0, 1\}$ <sup>1</sup>. The ROI is designated as the region of  $A$  corresponding to the non-zero elements of  $S$ .

The set of acquisition and reconstruction vectors are identified through explicit formulation and minimization of the cost function

$$\mathcal{J} = \|S \circ (A - AXL^H)\|_F^2, \quad (4.1)$$

where  $A$  and  $S$  are of size  $M \times N$ , and  $X$  and  $L$  are of size  $N \times r$ . The  $\circ$  operator denotes an element-by-element (Hadamard, or Schur) product. For an arbitrary matrix  $B$ , the Frobenius norm is defined as  $\|B\|_F^2 = \sum_{i,j} |b_{ij}|^2$ . We assume that  $A$  and any principal minor of  $A$  are full rank.

This cost function immediately suggests two problems which could be posed. On the one hand one can set an error tolerance level and seek a minimal  $r$  such that some  $X$  and  $L$  exist which produce a cost not in excess of that value. We term this the *minimal order* problem and discuss it in Section 4.2. Alternatively, we can fix  $r$  and seek an  $X$  and  $L$  which minimize  $\mathcal{J}$ . Section 4.3 is devoted to the analysis and solution of this *minimal error* formulation.

## 4.2 Minimal order problem

It turns out that the general case of the minimal order problem is quite intractable for mathematically precise reasons. To understand why, consider the simpler problem where we ask only for some  $Q \equiv XL^H$  such that the cost is zero. We ignore for the moment the requirement that  $Q$  be factorable into the  $XL^H$  form, with  $X$  and  $L$  of column width  $r$ , and seek only the individual elements of  $Q$  itself. This formulation belongs to a class of *matrix completion* problems [8, 21, 29, 20].

The best known matrix completion problems in signal processing involve maximum entropy

---

<sup>1</sup>Although selection matrices with binary elements are used here, the results can be extended to weighted selection matrices by using selection elements in the range  $0 \leq s_{ij} \leq 1$ .

extensions of autocorrelation sequences in which case the matrices possess a Toeplitz structure. Other common problems approach the completion of partially specified Hadamard or symmetric matrices. Solutions to these problems typically make deep use of the intended structural properties of the completed matrix,  $\hat{A}$ . The more general problem of choosing an unstructured  $Q$  (with or without the factorization constraint) and requiring the cost to be less than some non-zero threshold is much more complex. Other than its known usefulness in extending autocorrelation matrices for spectral estimation [32], no strong results for non-zero costs have been obtained to date.

Despite the difficulty in determining a solution to the general minimal order problem, we have found that there are cases with significant structure that allow us to say a bit more. We present two below which provide some useful insight and results which we use in our approach to the alternate, more tractable, minimal error formulation described in Section 4.3.

#### 4.2.1 Rectangular ROI, arbitrary error threshold

The first case of interest is when the ROI is rectangular in shape. In this case the optimal solution to the fixed error problem can be found from the SVD of the sub-matrix chosen by  $S$ . To begin, let us assume that  $S$  takes the form

$$S = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (4.2)$$

with  $\mathbf{1}$  the matrix of all ones. If the rectangular ROI is not located in the upper left corner, row and column permutations can be performed to arrive at the structure in (4.2). It is easily shown that  $\mathcal{J}$  is permutation invariant and no change in the cost results from these operations. Let  $A_{11}$  be the upper left block of  $S \circ A$  and define  $r_{11}$  to be the number of rows in  $A_{11}$ . With these definitions we have

**Theorem 1** *For rectangular ROIs, the solution to the minimal order, fixed error problem is given by the smallest  $r$  such that*

$$\sum_{i=r+1}^{r_{11}} \sigma_i^2 \leq \epsilon$$

where  $\epsilon$  is the error level, and  $\sigma_i$  is the  $i^{\text{th}}$  singular value of  $A_{11}$  with  $\sigma_1 > \sigma_2 > \dots > \sigma_{r_{11}}$ . Furthermore, the optimal  $X$  and  $L$  matrices for this solution can be obtained from the singular vectors of  $A_{11}$ .

*Proof:* Selection matrices with a rectangular ROI can always be permuted to the form of (4.2). Such matrices can be described by an outer product of two vectors,  $S = s_1 s_2^T$ . If the non-zero sub-block  $\mathbf{1}$  is of size  $m \times n$ , then  $s_1$  and  $s_2$  are vectors with  $m$  and  $n$  leading ones and  $(M - m)$  and  $(N - n)$  trailing zeros, respectively. As shown by Horn and Johnson in [18, p.304], a Hadamard product involving such a matrix can be rewritten as a conventional matrix product containing two diagonal matrices. Thus for matrices with a rectangular ROI, the cost function can be written as

$$\mathcal{J} = \|(s_1 s_2^T) \circ (A - AXL^H)\|_F^2 \quad (4.3)$$

$$= \|D_1(A - AXL^H)D_2\|_F^2 \quad (4.4)$$

where  $D_1$  and  $D_2$  are diagonal matrices with  $s_1$  and  $s_2$  along their respective diagonals. This can further be simplified to

$$\mathcal{J} = \|D_1 A D_2 - ZW^H\|_F^2 \quad (4.5)$$

where  $Z = D_1 A X$  and  $W = D_2 L$ . The optimal solution for  $W$  and  $Z$  can be found through the SVD of  $D_1 A D_2 = (S \circ A)$ . The structure of the optimal solution is

$$Z = D_1 A X = \begin{bmatrix} Z_1 \\ \mathbf{0} \end{bmatrix} \quad (4.6)$$

$$W^H = L^H D_2 = \begin{bmatrix} L_1^H & \mathbf{0} \end{bmatrix} \quad (4.7)$$

$$D_1 A D_2 = \begin{bmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (4.8)$$

Let the singular value decomposition of the rectangular sub-matrix  $A_{11}$  be  $A_{11} = U_1 \Sigma_1 V_1^H$ . The error at a given approximation  $r$  is therefore the sum of the discarded singular values, or equiva-

lently

$$\mathcal{J} = \sum_{i=r+1}^{r_{11}} \sigma_i^2$$

For the approximation to be less than a given error threshold  $\epsilon$ , one need only choose  $r$  such that

$$\sum_{i=r+1}^{r_{11}} \sigma_i^2 < \epsilon. \quad \blacksquare$$

Returning now to choose an optimal  $X$  and  $L^H$ , one may use the SVD decomposition to find

$$X = (D_1 A)^\dagger \begin{bmatrix} U_1 \Sigma_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} V_1 \\ \mathbf{0} \end{bmatrix} \text{ and } L^H = [V_1^H \ \mathbf{0}]. \quad (4.9)$$

This theorem has a number of interesting consequences. First, if the ROI is rectangular, then the SVD of the ROI (rather than the whole image) will in fact provide an optimal solution to our problem. More importantly, if the ROI is not rectangular, then the SVD of the smallest rectangle covering the ROI represents a sub-optimal solution and provides an upper bound on the error of the optimal solution to the underlying, arbitrary ROI problem. We use this observation in Section 4.3.3 to guide the determination of an appropriate order for the minimum error problem.

#### 4.2.2 Arbitrarily specified ROI, zero error

This second case concerns arbitrary ROIs and a fixed error of zero. Here we present a sub-optimal parameterization of  $(X, L)$  which provides the following sufficient condition for an order  $r$  solution to satisfy the zero error requirement:

**Theorem 2** *For a given selection matrix  $S$ , an order  $r$  solution of the form*

$$X = \begin{bmatrix} I_r \\ \mathbf{0} \end{bmatrix}, \quad L = \begin{bmatrix} I_r \\ Q_{12}^H \end{bmatrix} \text{ or } Q_r = XL^H = \begin{bmatrix} I_r & Q_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (4.10)$$

*will give zero error if  $\sum_i s_{ij} \leq r$  for each column  $j$  of  $S$  such that  $j > r$ . Here  $I_r$  is the  $r \times r$  identity matrix, and  $Q_{12}$  is an  $r \times (N - r)$  sub-matrix of free parameters. The minimum  $r$  for this form is found by permuting  $S$  such that its columns contain a non-increasing number of non-zero elements.*



*Proof:* This theorem is shown true by considering that for  $\mathcal{J} = 0$ , the following equation must hold,

$$S \circ A = S \circ (AXL^H) = S \circ (AQ_r). \quad (4.11)$$

From this, one can recognize that all columns in the  $Q_r$  formulation may be treated independently. The first  $r$  columns of the  $Q_r$  parameterization contain the identity matrix in the upper sub-block, and zeros elsewhere. Thus, the first  $r$  columns of the approximation  $(AQ_r)$  will be identical to the first  $r$  columns of  $A$ , satisfying (4.11) for those columns.

For the remaining columns, indexed from  $(r + 1)$  to  $N$ , each of the column equations can be rewritten as a system of equations with row size dependent on the number of non-zero elements in the  $j^{th}$  column of  $S$ . If one constructs the vector  $\alpha_j$  to contain the index values of the non-zero elements in the  $j^{th}$  column of  $S$ , then this column system may be written as

$$A(\alpha_j, j) = A(\alpha_j, 1 : r) q_j. \quad (4.12)$$

Here  $q_j$  is a length  $r$  vector containing the free parameters of the  $j^{th}$  column of  $Q_r$ . The vector  $A(\alpha_j, j)$  is composed of elements from the  $j^{th}$  column of  $A$  as specified  $\alpha_j$ . The matrix  $A(\alpha_j, 1 : r)$  is composed by taking certain rows as specified by  $\alpha_j$  from the first  $r$  columns of the original  $A$  matrix.

The number of rows in each column system depends on the number of ones in the  $j^{th}$  column of  $S$ . If  $\sum_i s_{ij}$  is greater than  $r$ , then the system is over-determined and the system can only be solved in an approximate sense. However, if  $\sum_i s_{ij}$  is less than or equal to  $r$ , then the system is under- or exactly-determined, and with our previous assumptions on  $A$ , an exact solution exists. Thus, if  $r$  is chosen such that  $\sum_i s_{ij} \leq r$  for each of the columns  $\{j; j > r\}$ , then none of the column systems will be over-determined.

Under this condition, collecting each of these column systems together, the order  $r$  solution of the form given in (4.10) has a sufficient number of free parameters to ensure that  $\mathcal{J} = 0$ . If we permute  $S$  such that the columns contain a non-increasing number of non-zero elements, we will

find the minimum value of  $r$  in the above expressions.  $\blacksquare$

We note here that other zero-error solutions may exist for a given problem, possibly with order less than the order of the  $Q_r$  solution. Thus, Theorem 2 provides an upper bound on the minimum order needed for a zero error solution. This result is used in the algorithm initialization discussion, presented in Section 4.3.2, and verified in Section 4.4.

To summarize, we have presented two results for the minimal order problem. On the one hand, we have a full solution to the problem for a rectangular region and arbitrary error. This result also provides an upper bound on the smallest  $r$  required for a given error threshold and arbitrary ROI. Second, we have a sufficient condition for a size  $r$  solution to the general ROI, zero error problem. This latter result provides an upper bound on the minimum order required to meet a zero error condition. Given that non-zero error solutions require fewer vectors than zero error solutions, this latter result also provides an upper bound on the order needed to meet any error threshold.

### 4.3 Minimal error, fixed order problem

Given the restrictive nature of the results in the previous section, we now present an alternate formulation to the problem of choosing  $X$  and  $L$ . Specifically, rather than fixing the error level and minimizing  $r$ , we fix  $r$  and find some  $X$  and  $L$  that provide minimum error. Formally, with the cost function given in (4.1) we seek a solution to

$$(X_{\text{opt}}, L_{\text{opt}}) = \underset{X, L}{\operatorname{argmin}} \mathcal{J}(X, L) = \underset{X, L}{\operatorname{argmin}} \|S \circ (A - AXL^H)\|_F^2 \quad (4.13)$$

for a given number of columns  $r$  in  $X$  and  $L$ .

We note that for any given  $r$ , the  $(X, L)$  pair that minimize  $\mathcal{J}$  are not unique. Any solution can be modified by an invertible matrix  $Z$  of appropriate size via  $XL^H = (XZ)(Z^{-1}L^H) = X_1L_1^H$ . In principle, one could think of either seeking an alternate parameterization of the problem which yields a unique solution or of using the extra degrees of freedom in  $R$  to achieve other design objectives for  $X$  and  $L$ . All we desire here are *some*  $X$  and  $L$  which minimize  $\mathcal{J}$ .

Because (4.13) is quartic in the elements of  $X$  and  $L$ , a minimum of  $\mathcal{J}$  cannot in general be determined in closed form. Thus we pursue a numerical solution to the optimization problem. To start, we observe that all minima of  $\mathcal{J}$  must satisfy both of the following equations:  $\partial\mathcal{J}/\partial X = 0$ , and  $\partial\mathcal{J}/\partial L = 0$ . Computing these partial derivatives gives

$$\frac{\partial\mathcal{J}}{\partial X} = -2A^H [S \circ S \circ (A - AXL^H)] L \quad (4.14)$$

$$\frac{\partial\mathcal{J}}{\partial L} = -2 [S \circ S \circ (A - AXL^H)]^H AX. \quad (4.15)$$

To determine an  $X$  and  $L$  satisfying (4.14) and (4.15) we employ a variant of the Cyclic Coordinate Descent (CCD) algorithm described by Luenberger in [28].

#### 4.3.1 CCD algorithm

The CCD algorithm alternately solves each of the two gradient equations, (4.14) and (4.15), once in each iteration. For each iteration step, first  $X$  is held fixed, and  $L$  is found such that  $\partial\mathcal{J}/\partial L = 0$ . To complete the iteration,  $L$  is held fixed, and a corresponding  $X$  is found such that  $\partial\mathcal{J}/\partial X = 0$ . Setting each of the gradient equations above, (4.14) and (4.15), equal to zero gives

$$A^H (S \circ S \circ A) L = A^H [S \circ S \circ (AXL^H)] L \quad (4.16)$$

$$[S \circ S \circ A]^H (AX) = [S \circ S \circ (AXL^H)]^H (AX). \quad (4.17)$$

These equations can be manipulated to yield a system of equations with the form  $Ba = c$  through the  $\text{vec}\{\}$  operator, which stacks the columns of a matrix into a column vector, and the Kronecker product,  $\otimes$  [18]. The vectorized versions of (4.16) and (4.17) are given in (4.18) and (4.19).

$$\text{vec}\{A^H (S \circ S \circ A) L\} = [(L^H \otimes A^H) \text{diag}\{\text{vec}\{S \circ S\}\} (L \otimes A)] \text{vec}\{X\} \quad (4.18)$$

$$\text{vec}\{[S \circ S \circ A]^H (AX)\} = [((AX)^H \otimes I_N) \text{diag}\{\text{vec}\{S^H \circ S^H\}\} ((AX) \otimes I_N)] \text{vec}\{L\} \quad (4.19)$$

Both of these vectorized systems contain a singular matrix of the form  $M^H \text{diag}\{\text{vec}\{S \circ S\}\}M$ . The elements of  $S$  appear as a diagonal matrix embedded in the middle of the matrix product. For region of interest problems, some elements of  $S$  will be zero, thereby causing the overall matrix to be rank deficient even if  $A$  is square and full rank. We use the Moore-Penrose pseudo-inverse (A.8) [17] to solve these systems at each iteration.

Although straightforward, a direct implementation of the coordinate descent algorithm is quite computationally intensive. Equations (4.18) and (4.19) both contain a matrix of size  $(rN) \times (rN)$  that must be solved via a pseudo-inverse at each iteration. However, as described in Appendix A.3, there exists a significant level of structure in these matrices that can be exploited to reduce the computational requirements for finding the system solution.

### 4.3.2 CCD algorithm initialization

The CCD method described in Section 4.3.1 converges to a local minimum of the cost function. Because many local minima may exist on the cost surface, convergence to a “good” minimum is dependent on the initialization point of the algorithm. After experimenting with a number of initialization heuristics, we found an approach that performed particularly well based on the  $X$  and  $L$  parameterization given in (4.10) of Section 4.2.2.

Specifically, substituting (4.10) into (4.13) leaves a *linear* least squares problem for determining  $Q_{12}$ . To determine  $Q_{12}$ , we solve a set of normal equations whose structure is similar to that of (4.17) and (4.19). We then form the matrix

$$Q_r = \begin{bmatrix} I_r & Q_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

and compute its SVD,  $Q_r = U\Sigma V^H$ . The CCD algorithm is initialized with  $X = U\Sigma$  and  $L = V$ .

### 4.3.3 Choice of approximation order

Solution of the minimum error problem requires a prior specification of the approximation order  $r$ , i.e., the number of vectors in  $X$  and  $L$ . Here we concentrate on selecting the order based on upper bounds of the resulting error. Given the discussion in Section 4.2, it is not surprising that we have two types of bounds: one error bound based on an SVD argument for the case where we allow the cost to take on some finite, non-zero value; and one order bound based on the restricted forms of  $X$  and  $L$  in (4.10).

For those cases where a non-zero error is acceptable, one may use an SVD of the smallest rectangle covering the ROI to find an upper bound on the ROI reconstruction error at a given order. As shown in Theorem 1, the SVD is optimal for reconstruction of a rectangular ROI. If the SVD solution at a given order can provide an acceptable approximation error, we can guarantee that the error resulting from the localized ROI  $(X_{\text{opt}}, L_{\text{opt}})$  solution to (4.13) will be no larger.

To verify this claim, one need only consider the following. Let  $(X', L')$  be (the optimal) vectors obtained from the SVD of the smallest rectangle covering the ROI. Let  $(X_{\text{opt}}, L_{\text{opt}})$  be the solution to (4.13) for reconstructing the ROI. In the case that  $(X_{\text{opt}}, L_{\text{opt}}) = (X', L')$ , the reconstruction error for each set will be equal as well. We have shown that in general, the  $(X', L')$  solution does not minimize  $\mathcal{J}$  for an arbitrarily shaped ROI while  $(X_{\text{opt}}, L_{\text{opt}})$  does minimize  $\mathcal{J}$ . Thus, one can generally expect the  $(X_{\text{opt}}, L_{\text{opt}})$  solution to give an ROI reconstruction with lower error than  $(X', L')$ , and certainly the error will be no larger. Of course the CCD algorithm presented in this section to find  $(X_{\text{opt}}, L_{\text{opt}})$  can only converge to a local minimum, depending on the given initialization point. However, as illustrated in the examples below, our experience has been that the CCD solution provides significantly less reconstruction error than the SVD solution,  $(X', L')$ .

For arbitrarily shaped regions of interest, Theorem 2 of Section 4.2.2 provides an upper bound on the minimum order for zero error, which we denote as  $r_0$ . According to this theorem, after permuting  $S$  and then comparing each successive column index to the number of non-zero elements

in that column, the upper bound,  $r_u$ , is the smallest  $r$  that satisfies the inequality  $\sum_i s_{ij} \leq r$  for all  $j > r$ .

## 4.4 Examples

This section presents results from the application of the CCD method to a few simulated examples and experimental lab results. For the images in this section, the region of interest is shown with a standard intensity map, while the region outside the ROI is shown with an inverse intensity map. That is, outside the ROI, pixels of high intensity are shown darker than pixels of low intensity.

### 4.4.1 Simulation results

This simulation example illustrates a comparison between the Cyclic Coordinate Descent (CCD) method and two other low order approximation techniques currently used in MR imaging: Singular Value Decomposition (SVD) and Low-order Fourier (LoF). For the LoF reconstructions, only the  $r$  lowest spatial frequency components of the smallest sub-matrix of  $A$  containing the ROI were used in the reconstruction. For the SVD reconstruction, only the right singular vectors of the smallest sub-matrix of  $A$  containing the ROI corresponding to the  $r$  largest singular values were used. The CCD reconstruction vectors were found as described in Section 4.3.1.

Figure 4.1 shows an MRI scan of a human head along the sagittal plane. The ROI selection matrix is contained within a  $94 \times 54$  pixel rectangle. Thus to achieve zero error using either the SVD or LoF techniques, 54 input vectors would be needed. Given the sparseness of the ROI, we expect the order of the zero error CCD solution,  $r_0$ , to be much lower than this. The permuted selection matrix used to determine the upper bound  $r_u$  is shown in the right panel of Figure 4.2. The non-zero element count for each column and a marker for the upper bound  $r_u$  is shown in the right panel. The upper bound on  $r_0$  is determined as per Section 4.3.3 and is found to be  $r_u = 43$ . We note that if the upper bound is tight, using 43 input vectors to re-scan the ROI will still give

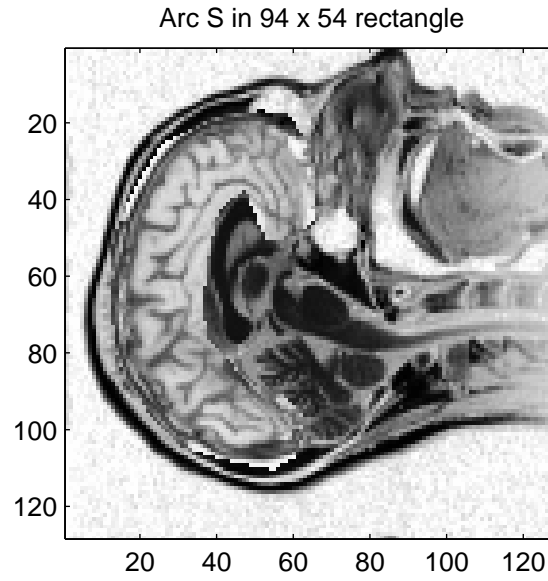


Figure 4.1: Original MR Image and ROI for static simulation example. The region outside the ROI is indicated with an inverse intensity map.

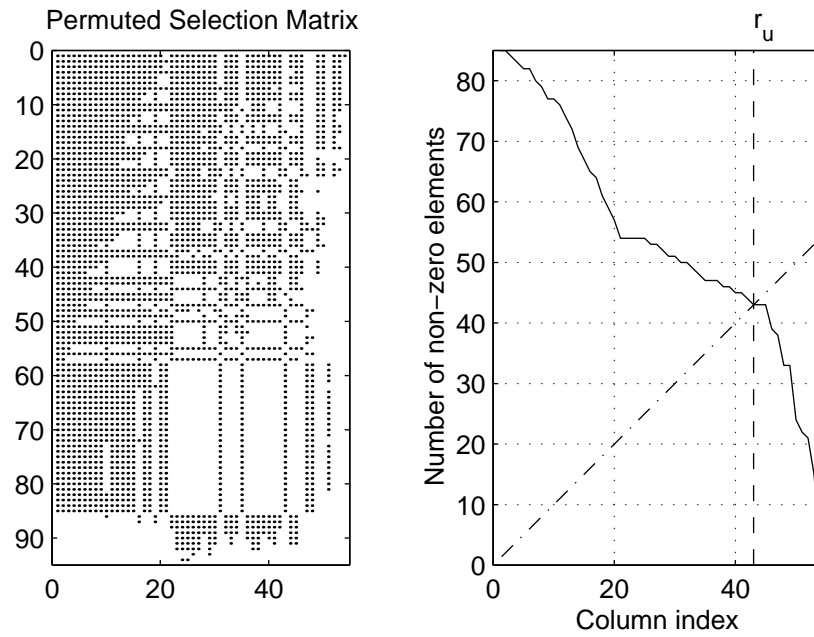


Figure 4.2: Permuted selection matrix for Figure 4.1 and geometric determination of  $r_u$ .

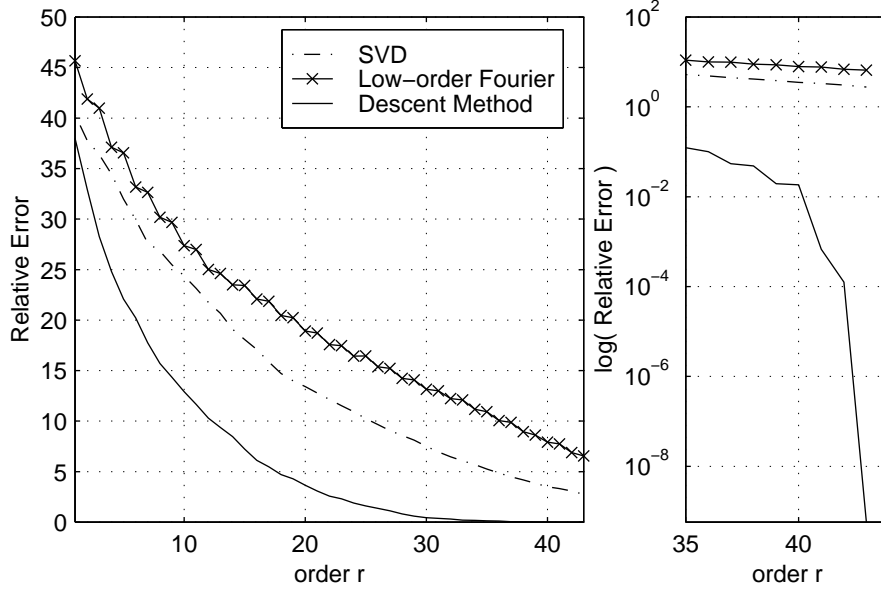


Figure 4.3: Relative error comparison of SVD, Low-order Fourier, and Cyclic Coordinate Descent solutions for Figure 4.1 ROI. Right panel shows logarithmic scale for better detail near zero error per pixel.

a decrease in the acquisition time of about 20% compared to the SVD and LoF techniques, with zero error in the ROI.

Figure 4.3 compares the relative error (A.7) for the three different methods of low order approximation over a range of solution orders. The right panel shows the upper range of orders with the average error per pixel value plotted on a log scale to show greater detail near zero error. The figure shows that at a given order, the CCD solution provides lower reconstruction error than either the LoF or the SVD method. Furthermore, Figure 4.3 shows that for a given error tolerance, a CCD reconstruction of the ROI is available at a much lower order than either the SVD or Fourier approximation methods provide. For instance, if the number of input vectors is fixed at 10, the CCD solution has an average pixel error that is one half that given by the SVD. Conversely, for a given error per pixel of 10, the ROI can be acquired in less than half the time using the CCD method. We found similar results for the many image examples we examined with a general ROI specified.



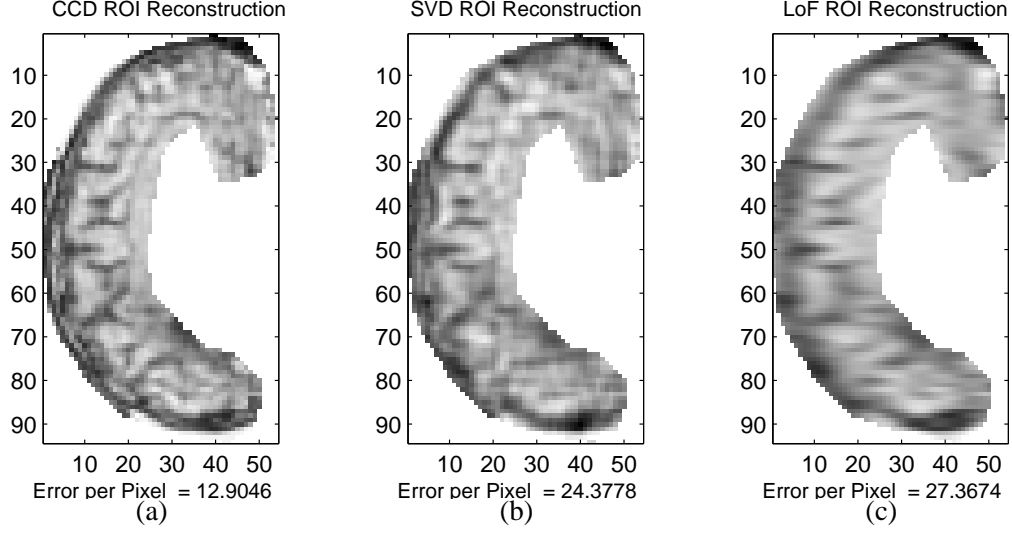


Figure 4.4: Comparison of order  $r = 10$  ROI reconstructions for (a) Cyclic Coordinate Descent  $[\hat{A}_{CCD}]$ , (b) SVD  $[\hat{A}_{SVD}]$ , and (c) Low-order Fourier  $[\hat{A}_{LoF}]$  methods

Comparison of the three ROI reconstruction methods, (SVD, LoF, and CCD), are given below for two specific orders,  $r = 10$  and  $r = 25$ .

Figure 4.4 shows the order 10 ROI reconstructions. The absolute difference in pixel values for the same methods and order are shown in Figure 4.5. It is clear from both figures that the solution found by the CCD method has significantly less pixel error than either the LoF or SVD methods. The CCD solution also shows a more even distribution of the error across the ROI, and greater structural information in the ROI than either of the global orthogonal approximation methods.

The order 25 reconstructions are shown in Figure 4.6, with the absolute error illustrated in Figure 4.7. While all three approximations now show structural detail in the ROI, there is still an order of magnitude difference in the error per pixel measure. This is confirmed visually in the absolute difference illustrations shown in Figure 4.7. Here, negligible error is shown for the CCD solution, while significant error still occurs in the other two.

As seen in this example, the CCD solution is able to provide image acquisition and reconstruction vectors that are tailored to represent local information in the ROI. This method needs

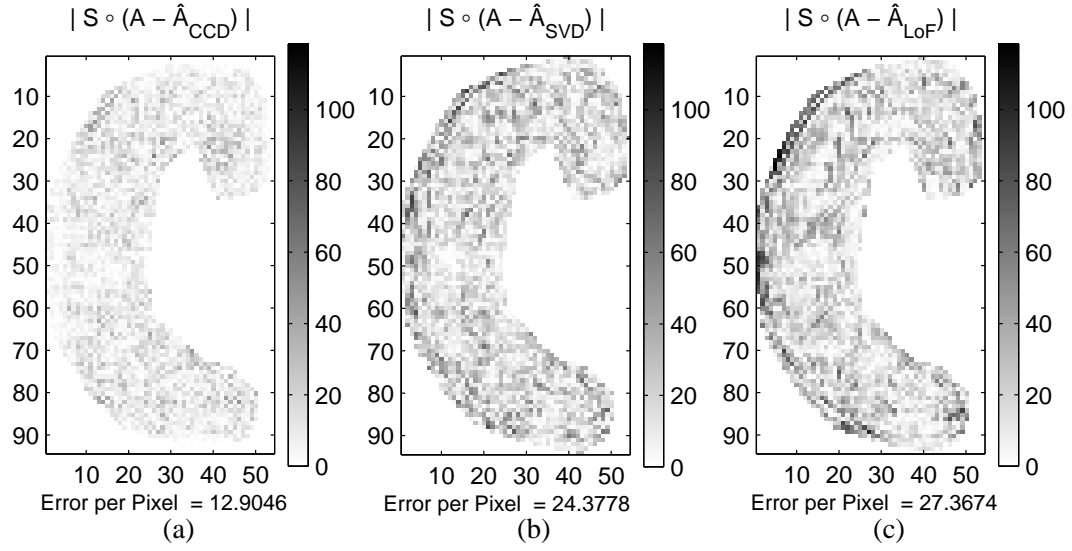


Figure 4.5: Pixel value difference comparisons of order  $r = 10$  ROI reconstructions for (a) Cyclic Coordinate Descent  $|A - \hat{A}_{CCD}|$ , (b) SVD  $|A - \hat{A}_{SVD}|$ , and (c) Low-order Fourier  $|A - \hat{A}_{LoF}|$  methods

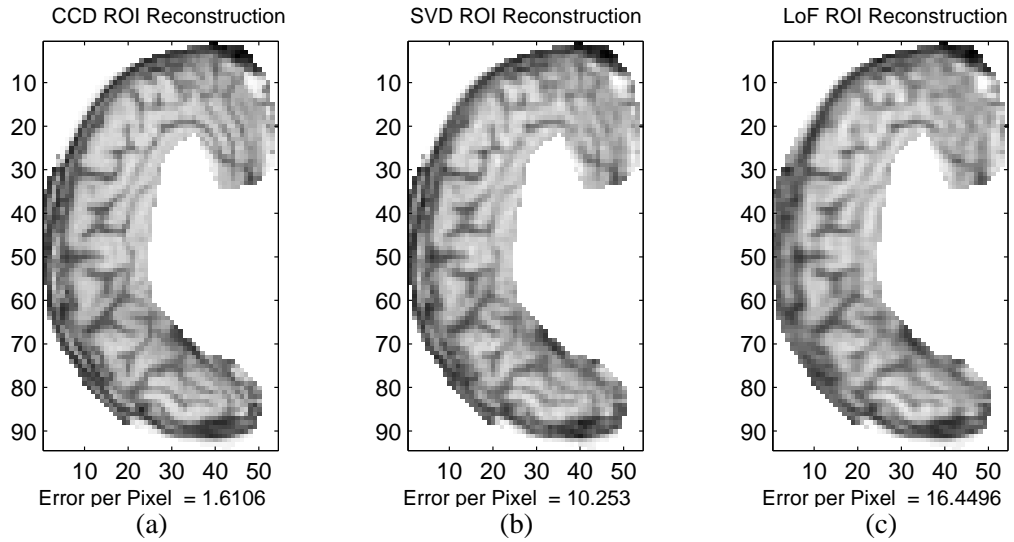


Figure 4.6: Comparison of order  $r = 25$  ROI reconstructions for (a) Cyclic Coordinate Descent  $[\hat{A}_{CCD}]$ , (b) SVD  $[\hat{A}_{SVD}]$ , and (c) Low-order Fourier  $[\hat{A}_{LoF}]$  methods

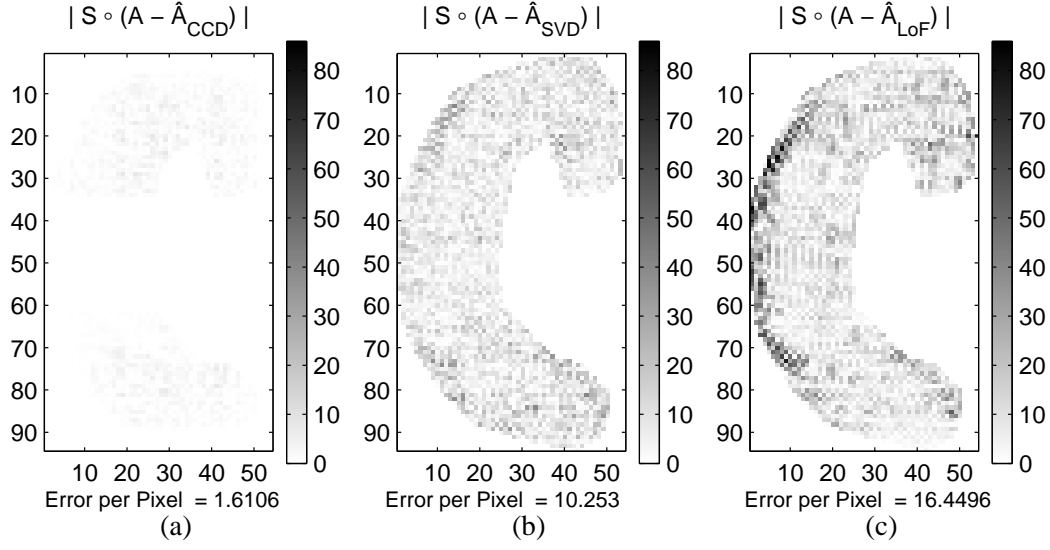


Figure 4.7: Pixel value difference comparisons of order  $r = 25$  ROI reconstructions for (a) Cyclic Coordinate Descent  $|A - \hat{A}_{\text{CCD}}|$ , (b) SVD  $|A - \hat{A}_{\text{SVD}}|$ , and (c) Low-order Fourier  $|A - \hat{A}_{\text{LoF}}|$  methods

significantly fewer vectors to reconstruct the ROI image with quality comparable to the SVD method. These results suggest a significant decrease in acquisition time savings is possible for MRI acquisitions using this method.

#### 4.4.2 Laboratory results

While the simulation results illustrated in Section 4.4.1 above show promise, the MRI lab experiments attempting to use the CCD method have been less than satisfactory. Noise in the laboratory environment is non-negligible, and experimental results indicate that the CCD method is very sensitive to noise.

This section shows reconstructions of a phantom slice for two ROIs: one completely covering the phantom, and one completely in the interior of the phantom. The experiments were performed by placing a phantom in a GE Signa MRI scanner which is rated with 1.5 Tesla coil. The image data was acquired using a modified spin-echo sequence that allows for non-Fourier imaging via rf-encoding. First the full  $k$ -space data for the central slice was acquired and the spatial domain

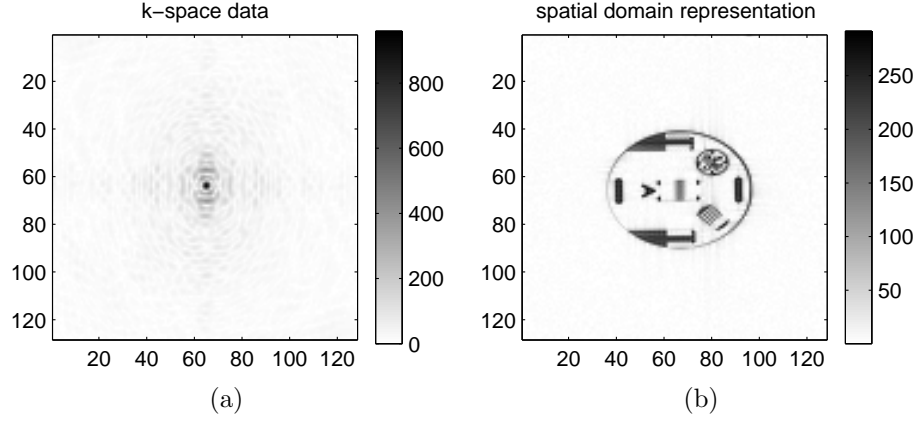


Figure 4.8: Original phantom image for static laboratory example. (a)  $k$ -space data (b) spatial domain image

image was formed via (3.9) using the unitary Fourier transform matrix  $F$  as given in 3.8. Figure 4.8 shows the reference data in both the  $k$ -space and spatial domain. This data was used to guide the input/reconstruction vector choice for the ROI acquisitions that follow.

A selection matrix  $S$ , shown in Figure 4.9, was created to completely cover the phantom in the image slice. The image and selection matrices were then passed to the CCD algorithm, which produced a set of vectors  $X$  and  $L$  designed to efficiently acquire/reconstruct the ROI for subsequent scans.

The first set of figures below show reconstructions for the covering ROI using three different choices of inputs. For a given ROI acquisition/reconstruction vector set  $XL^H$ , one can distribute the power between  $X$  and  $L$  with an invertible matrix  $Z$  via  $XZZ^{-1}L^H$ . The experiments below focus on the distribution of the singular values of  $XL^H$ . Given the SVD of  $XL^H = U\Sigma V^H$ , the experiment shown below is for the three cases  $X = U$ ,  $X = U\Sigma^{1/2}$ , and  $X = U\Sigma$ .

The figures correspond to the following three cases:

Case	Input	ROI Reconstruction	Figure
1	$U$	$Y\Sigma L^H = (AU + N)\Sigma L^H$	4.10
2	$U\Sigma^{1/2}$	$Y\Sigma^{1/2}L^H = (AU\Sigma^{1/2} + N)\Sigma^{1/2}L^H$	4.11
3	$U\Sigma$	$Y\Sigma L^H = (AU\Sigma + N)L^H$	4.12

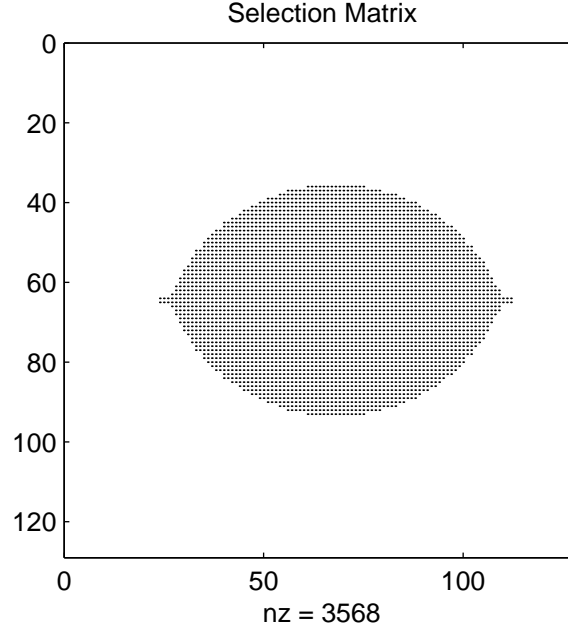


Figure 4.9: Covering ROI for lab experiments

A second experiment was performed with the ROI covering only an interior portion of the phantom. While the reconstructions are somewhat better, they still contain substantially more image distortion than predicted in simulation. The interior ROI is shown in Figure 4.13. Figure 4.14 shows the expected and actual reconstructions, respectively. As seen in the right side of Figure 4.14, the image reconstructed from the experimental data shows a portion of the image accurately, but there is still significant error close to the edges of the phantom.

For both the covering and interior ROI examples, the reconstructions are clearly not satisfactory. Preliminary analysis shows this is primarily a consequence of noise in the measured data. Approaches to address this problem are presented in more detail in Chapter 6.

## 4.5 Summary of the static problem

This section showed that there exists a set of acquisition/reconstruction vectors to efficiently represent local signal information in a specified ROI of an image. This section also presented an

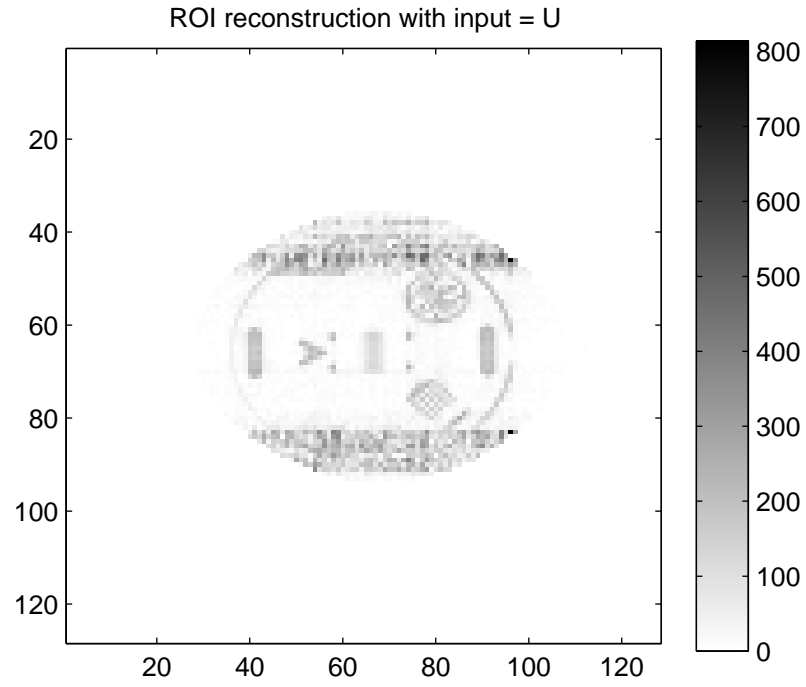


Figure 4.10: ROI reconstruction for  $X = U$ , the right singular vectors of  $XL^H$ .

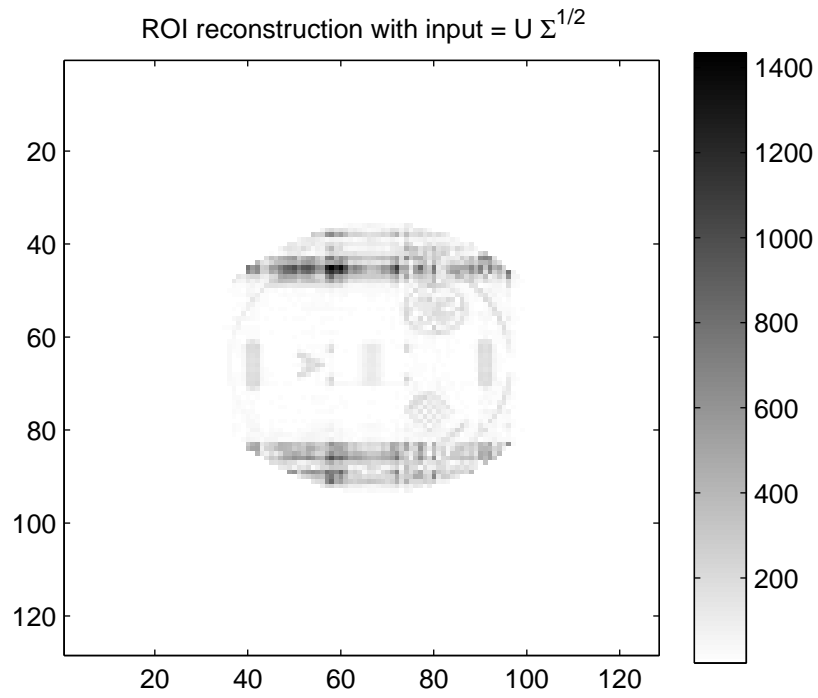


Figure 4.11: ROI reconstruction for  $X = U \Sigma^{1/2}$ , the scaled right singular vectors of  $XL^H$ .

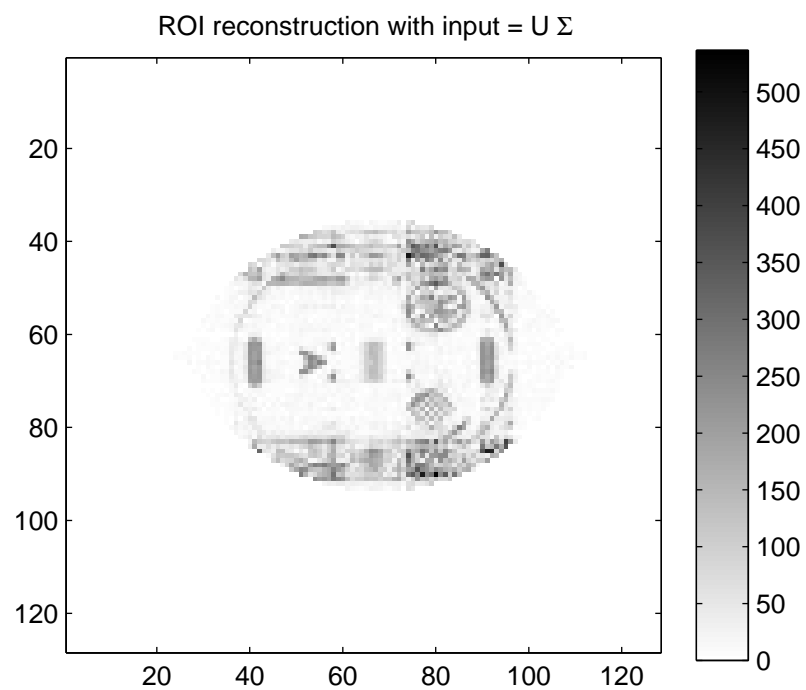


Figure 4.12: ROI reconstruction for  $X = U\Sigma$ , the scaled right singular vectors of  $XL^H$ .

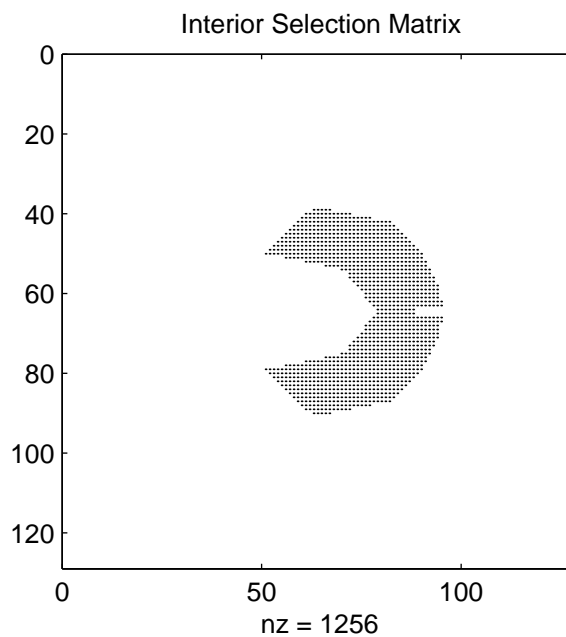
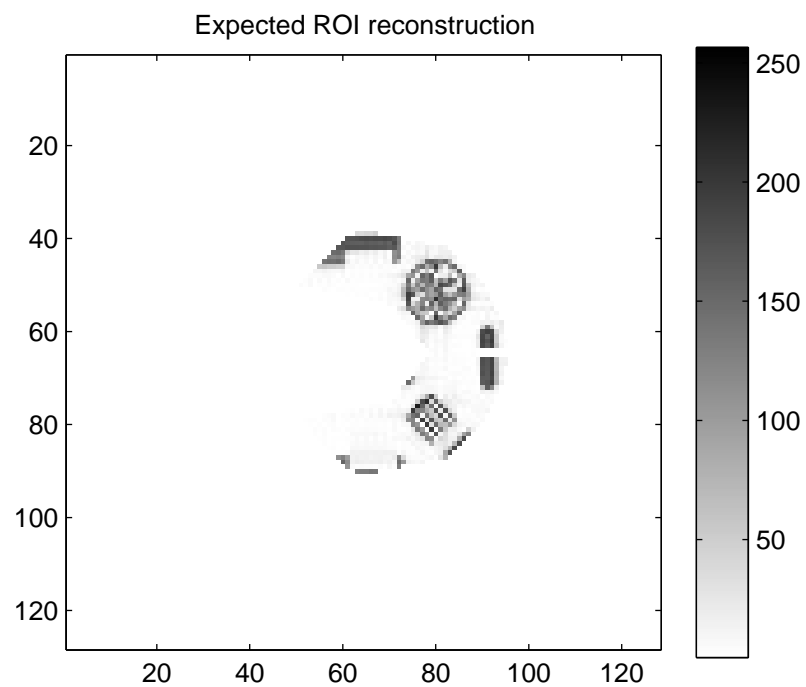
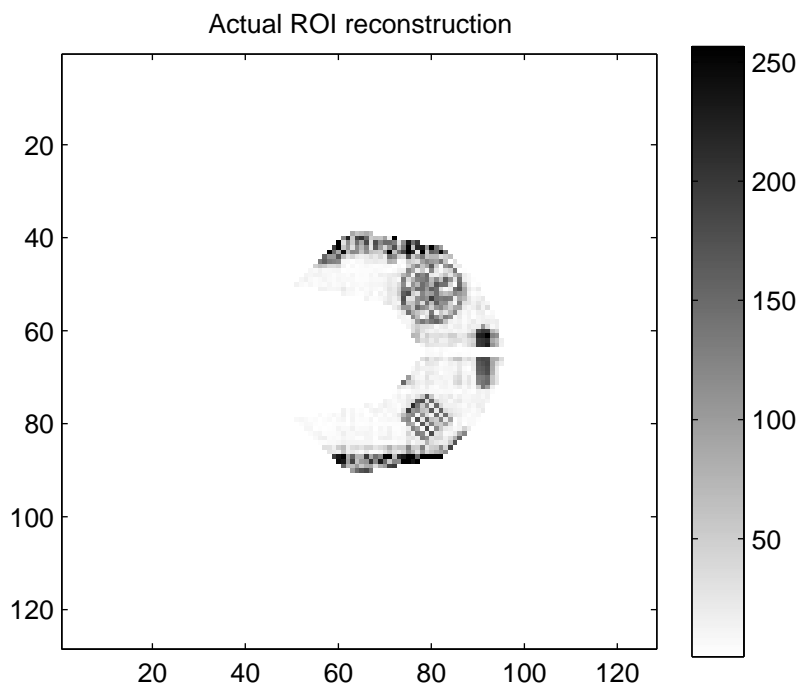


Figure 4.13: Interior ROI for lab result example



(a)



(b)

Figure 4.14: Reconstruction of interior ROI images



iterative algorithm using the Cyclic Coordinate Descent method to find such an image acquisition vector set. This method accurately reconstructs ROIs of a given image matrix in simulation. However, the practical use of this method in the lab is limited. We suspect other design criteria must be included in order to compensate for the noisy nature of experimental data. A strategy to design such criteria will be examined below in Chapter 6.

## Chapter 5

# Adaptive modeling of the dynamic MRI process

The main goal of low-order dynamic MRI is to reduce the sequence acquisition time through the application of efficient imaging methods. As shown in Chapter 3, approaching the problem from a subspace identification perspective (vis-à-vis the SVD) and applying model based estimation (vis-à-vis RIGR) can be advantageous. This section seeks to formalize and examine such methods, building primarily on the linear subspace method as proposed by Panych and Zientara, et. al., in [47, 36]. Based on this analysis, a general adaptive estimation framework and two new methods for choosing input vectors are introduced.

The major difference between the problem studied in this section and other dynamic estimation and tracking problems is that there is almost complete control of the system inputs to acquire and track dynamic MRI sequences. This results in a *doubly adaptive system*, i.e., both the image estimate and the input acquisition vectors must be determined at each point in the sequence. And much like the classic “Which came first, the chicken or the egg” paradox, the resulting quality of each half of the doubly adaptive system depends directly on the other.

This section details two distinct, yet coupled, dynamic sequence acquisition sections. First, a general adaptive framework is developed in Section 5.1. The focus there is on the traditional adaptive filter approach of minimizing  $\|Y_n - \hat{Y}_n\|_F^2$ , where  $Y_n$  is the output of the system to be identified and  $\hat{Y}_n$  is the estimated output of the adaptive system. By analyzing the image reconstruction and estimation process in this manner, we find that each of the low-order methods described in Chapter 3 are closely related. In fact, a general adaptive framework is presented which is a superset of the FK and SVD methods.

Second, methods to identify appropriate input vectors are presented in Section 5.2. While the image estimation discussion of Section 5.1 provides a theoretically optimal set of input vectors, finding vectors that produce high quality image estimates is in fact quite difficult. The reason is that the image estimates themselves are closely related to the input vectors used to acquire the data used in the image reconstruction. Thus, input vectors chosen using previous image estimates will be closely biased to the previous input vector set. To escape this bias, some cleverness must be introduced drawing from a clear understanding that each new input vector set must “look in new places.” Two methods that succeed at this are presented below in Section 5.2.

Finally, this section closes with a comprehensive comparison of the low-order imaging methods reviewed in Chapter 3 and the new methods presented in this section. The comparisons cover sequences simulating real dynamic MRI acquisitions and synthetic sequences designed to isolate features common to dynamic MRI. As shown in the examples of Section 5.3, the new methods perform quite well and are in fact quite suitable for dynamic MRI sequences that exhibit dramatic motion changes.

## 5.1 Construction of the image estimate

The goal of low order dynamic MRI is to estimate an image at time  $A_n$  using a limited number of input scans. The data available to achieve this includes all of the previous outputs up to

$Y_n = A_n X_n$ , and the corresponding input matrices,  $X_n$ . As discussed in the introductory section above, the focus in this section is on constructing estimates using “tall-and-skinny” orthogonal inputs  $X_n$  of size  $r \times N$ , with  $r < N$  and a single input/output pair of matrices,  $Y_n$  and  $X_n$ . We identify the image estimate through minimization of the difference between the measured output data and the predicted output data. Analytically, this is described by finding  $\hat{A}_n$  such that the cost function

$$\mathcal{J}_n = \|Y_n - \hat{A}_n X_n\|_F^2 = \|Y_n - \hat{A}_n X_n\|_F^2 \quad (5.1)$$

is minimized. This problem is underdetermined. This is due to the choice that the number of column vectors  $r$  in  $X_n$  and  $Y_n$  is less than  $N$ . Other possibilities exist, some of which are discussed in Chapter 6. In (5.1), the number of parameters available in the image estimate  $\hat{A}_n$  is larger than the number of constraints given by  $Y_n$ . Consequently, an infinite number of zero error solutions exist. The treatment that follows explores a few solutions for the underdetermined problem in (5.1). In the process, we show that the low-order reconstruction methods given in Chapter 3 are in fact all contained within the same common framework.

First of all, one could solve  $\mathcal{J}_n = 0$  directly. With no structural constraints on  $\hat{A}_n$ , this leads to the underdetermined system  $\hat{A}_n X_n = Y_n$ . One solution is

$$\hat{A}_n = Y_n X_n^H (X_n X_n^H)^\dagger. \quad (5.2)$$

If we constrain the columns of  $X_n$  to be orthonormal,  $X_n^H (X_n X_n^H)^\dagger = X_n^H$ , Equation (5.2) reduces to the *low-rank reconstruction* solution:

$$\hat{A}_n = Y_n X_n^H = A_n X_n X_n^H. \quad (5.3)$$

This low rank estimate was used by Panych, et. al., in their paper describing the SVD encoding method [34].

One could instead minimize (5.1) while incorporating information from a reference image. Traditionally, this reference image is obtained at the start of the sequence [43, 47, 36], or may be

constructed from a collection of similar images [6]. To incorporate reference image information in the image estimate, we model the image changes as  $A_n = A_0 + \alpha_n$ , and similarly  $\hat{A}_n = A_0 + \hat{\alpha}_n$ . Solving  $\mathcal{J}_n = 0$  with this model for  $\hat{A}_n$  gives

$$\begin{aligned} Y_n - \hat{A}_n X_n &= Y_n - (A_0 + \alpha_n) X_n &= 0 \\ \alpha_n X_n &= Y_n - A_0 X_n \\ \alpha_n &= (Y_n - A_0 X_n) X_n^H (X_n X_n^H)^\dagger \end{aligned}$$

Again, with the constraint that the columns of  $X_n$  are orthonormal, this leads to

$$\hat{A}_n = A_0 + Y_n X_n^H - A_0 X_n X_n^H = Y_n X_n^H + A_0 (I - X_n X_n^H) \quad (5.4)$$

Upon inspection, if the inputs  $X_n$  are constructed from the low frequency components of the Fourier basis, Equation (5.4) is recognized as the *keyhole reconstruction* method, as first proposed by Brummer and Van Vaals, et. al., [4, 43]. Furthermore, if  $A_0 = \mathbf{0}$ , then this solution is identical to the low rank reconstruction in (5.3).

A third possibility is to solve  $\mathcal{J}_n = 0$  while incorporating information from the most recent estimate,  $\hat{A}_n = \hat{A}_{n-1} + \beta_n$ . This leads to

$$\begin{aligned} Y_n - \hat{A}_n X_n &= Y_n - (\hat{A}_{n-1} + \beta_n) X_n &= 0 \\ \beta_n X_n &= Y_n - \hat{A}_{n-1} X_n \\ \beta_n &= (Y_n - \hat{A}_{n-1} X_n) X_n^H (X_n X_n^H)^\dagger. \end{aligned}$$

In this case, by imposing the orthogonality constraint on  $X_n$ , we find that minimizing (5.1) gives an *adaptive framework*:

$$\hat{A}_n = Y_n X_n^H + \hat{A}_{n-1} (I - X_n X_n^H). \quad (5.5)$$

Rewriting (5.5) as  $\hat{A}_n = \hat{A}_{n-1} + (Y_n - \hat{A}_{n-1} X_n) X_n^H$ , one can recognize a similarity between the adaptive framework and the least mean square (LMS) adaptive algorithm

$$\text{Filter output:} \quad y_n = \hat{\mathbf{w}}_n^H \mathbf{u}_n$$

$$\text{Estimation error:} \quad e_n = d_n - \hat{\mathbf{w}}_n^H \mathbf{u}_n$$

$$\text{Tap-weight adaptation:} \quad \hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu \mathbf{u}_n e_n^*$$

as given by Haykin [14, Chap. 9]. In the LMS algorithm,  $\mathbf{u}_n$  is the tap-input vector,  $d_n$  is the desired response at time  $n$ ,  $e_n$  is the estimation error, and  $\mu$  is the step-size parameter. Comparing the LMS algorithm with (5.5), we see the step-size parameter is  $\mu = 1$  and the error term  $e_n$  corresponds to  $\mathcal{J}_n$ . This value for the step-size parameter is not surprising. From Haykin [14, § 9.4], the LMS algorithm is convergent in the mean squared sense if  $\mu$  satisfies the condition  $0 < \mu < 2/\lambda_{\max}$  where  $\lambda_{\max}$  is the largest eigenvalue of the input correlation matrix — a measure of the input signal power. With  $X_n$  formed from orthonormal columns,  $\lambda_{\max} = 1$ , and a step-size parameter value of  $\mu = 1$  places (5.5) squarely in the stable convergence region.

For the purpose of comparison, consider (5.5) for the limiting case of a static input vector set,  $X_n = X \forall n$ , formed from orthonormal columns. Consider that in this case, the projection of the low-order reconstruction terms  $Y_k X^H$  onto the complementary subspace  $(I - X X^H)$  will result in  $Y_k X^H (I - X X^H) = 0$ . This leads to

$$\begin{aligned} \hat{A}_n &= Y_n X^H + \hat{A}_{n-1} (I - X X^H) \\ &= Y_n X^H + [Y_{n-1} X + \hat{A}_{n-2} (I - X X^H)] (I - X X^H) \\ &= Y_n X^H + \hat{A}_{n-2} (I - X X^H) \\ &\vdots \\ \hat{A}_n &= Y_n X^H + \hat{A}_0 (I - X X^H). \end{aligned}$$

Notice that the cancellation effect occurs all the way back to the original reference image  $A_0$ . There are no contributions from the intermediate images  $A_k$  in the estimate of  $\hat{A}_n$  for  $0 < k < n$ . Thus, with static orthogonal inputs, the adaptive framework solution for  $\hat{A}_n$  is fundamentally identical to the keyhole method described in (5.4). Because the FK and SVD methods are special cases of (5.5), we describe this solution as the general adaptive estimate framework.

Note that in each of these three cases,  $\mathcal{J}_n = \|Y_n - \hat{Y}_n\|_F^2 = 0$ . One can only determine which is the best solution by comparing the actual image  $A_n$  to the estimated image  $\hat{A}_n$ . This is done via the absolute error measure

$$\mathcal{E}_n = \|A_n - \hat{A}_n\|_F^2. \quad (5.6)$$

Equation	Reconstruction Method	$\hat{A}_n =$	$\mathcal{E}_n = (A_n - \hat{A}_n) =$
(5.3)	Low-rank	$A_n X_n X_n^H$	$A_n (I - X X^H)$
(5.4)	Keyhole	$A_n X_n X_n^H + A_0 (I - X_n X_n^H)$	$(A_n - A_0) (I - X X^H)$
(5.5)	Adaptive Framework	$A_n X_n X_n^H + \hat{A}_n (I - X_n X_n^H)$	$(A_n - \hat{A}_{n-1}) (I - X_n X_n^H)$

Table 5.1: Image reconstruction method summary

The absolute estimate error in each image reconstruction method discussed above is shown in Table 5.1. To minimize  $\mathcal{E}_n$ , this table gives an indication of the best input vectors to use in each case: the right singular vectors (rSV) of the next image for (5.3); the rSV of the difference between the next image and the reference image for (5.4); and the rSV of the difference between the next image and the previous estimate for (5.5). However,  $A_n$  is not known ahead of time. Thus, the optimal input vector selection choice implied by Table 5.1 is a theoretical bound on the estimate quality for a given image reconstruction method.

With a framework for constructing image estimates now established, the next section examines the second half of the image estimation process: input vector identification.

## 5.2 Input vector identification

Low order MRI acquisition techniques became prominent in the mid 1990's. Fourier based methods include building an image estimate from a limited number of  $k$ -space lines [25, 13], and novel  $k$ -space sampling to reconstruct regular sub-regions of the image [31]. Fourier-based input sets have also been derived from a large aggregate set of similar images [7]. Non-Fourier methods include

the use of wavelets [15] and identifying a set of input vectors via the SVD of a full image acquired at the beginning of the sequence [47].

In each of these cases the input vectors used to acquire the MRI data are orthonormal. This assumption/constraint of orthonormal input vectors is continued here for two reasons. One, implementation of the acquisition vectors in the MRI scanner is aided by this constraint. The scanner inputs are limited to a fixed-point precision representation and are typically scaled such that the peak absolute value of the data is “1”. The use of orthonormal inputs ensures that the dynamic range of the input sequence is compatible with the scanner implementation. Second, from an analytical perspective, if  $N$  orthogonal inputs are applied to a static image in succession, one can ensure that the estimate of  $A$  of size  $M \times N$  equals the true image data. That is, if the input vectors span the full image space, a full scan of the image can be acquired using a non-Fourier basis set while scanning the same number of  $k$ -space lines as traditional Fourier acquisition methods. The easiest way to ensure a full span of the image space is to use orthonormal input vectors. Third, it greatly simplifies the analysis that follows.

In the acquisition methods cited above, one fixed set of input vectors is chosen to acquire the entire dynamic sequence and a low-order estimate of the image is formed by applying an image model to the output data. However, due to the dynamic nature of MRI sequences, a fixed set of basis vectors may not be best over an entire sequence. Moreover, an input vector that works well for one sequence may not work well for a different sequence. For this reason, methods to dynamically select input vectors over the course of the sequence are the focus of this section.

### 5.2.1 The subspace trap

Table 5.1 in Section 5.1 showed that the best choice of input vectors is dependent on the image to acquire, and is realizable only when the next true image is known. Since each image in the sequence  $A_n$  is not available ahead of time, one must select the input vectors based on image estimates. However, determining input vectors based solely on the estimate of prior images tends to bias the



new input vectors towards the previous inputs. In fact, analysis shows that many similar methods one could use to determine a new set of inputs result in  $X_n = X_{n-1}$ . We refer to this tendency as the *subspace trap* and it is examined in detail below.

The clearest example of the subspace trap occurs in the following case. Consider choosing a new set of input vectors by finding the right singular vectors of the current estimate. If the estimate method used is the low-rank reconstruction,  $\hat{A}_n = Y_n X_n^H = A_n X_n X_n^H$ , then the right singular vectors of the estimate will in fact always span the basis spanned by the input vector matrix  $X_n$ .

A similar conclusion is reached when using a keyhole style reconstruction,  $\hat{A}_n = A_n X_n X_n^H + A_0(I - X_n X_n^H)$ . Let the vectors for the first acquisition  $X_1$  be chosen from the right singular vectors of the reference image  $A_0$  at the start of the sequence. With this construction,  $X_1$  is chosen to minimize  $\|A_0(I - X_1 X_1^H)\|_F^2$ . So, unless the magnitude of the image information captured by the input,  $A_1 X_1 X_1^H$ , is on the order of the reference image noise,  $A_0(I - X_1 X_1^H)$ , the subspace spanned by  $X_1 X_1^H$  will tend to dominate the right singular vectors of  $\hat{A}_n$ . So although this case is not as limiting as the previous case, using the right singular vectors of  $\hat{A}_n$  again results in a subspace trap.

For the adaptive estimate framework given in (5.5), it was shown in Section 5.1 that the best input vectors were in fact found from the right singular vectors of the difference between the next true image and the estimate of the previous image. Because the true images  $A_n$  are unavailable, a reasonable choice might be to substitute estimates of these images. Consider then the case of finding the right singular vectors of the difference matrix  $(A_n - \hat{A}_{n-1})$ . At  $n = 1$ , the previous estimate is the reference image,  $\hat{A}_0 = A_0$ , which is presumed to be known. Let a discrete change occur in the image system,  $A_1 = A_0 + dA$ . The output vectors from the first acquisition are then

$$Y_1 = A_1 X_1 = (A_0 + dA) X_1,$$

and the adaptive framework estimate of  $A_1$  is

$$\hat{A}_1 = \hat{A}_0(I - X_1 X_1^H) + Y_1 X_1^H$$

$$= \hat{A}_0 + dAX_1X_1^H.$$

Substituting this estimate for  $A_1$  in the difference matrix, one finds

$$(\hat{A}_1 - \hat{A}_0) = (A_0 + dAX_1X_1^H - A_0) = dAX_1X_1^H$$

Clearly, the right singular vectors of this matrix will span the subspace described by  $X_1$ , regardless of the structure of  $dA$ . So by this method, the subspace spanned by the input vectors will stay static throughout the dynamic sequence — another subspace trap.

Finally, consider the case of determining new inputs by finding vectors  $X_{n+1}$  that minimize the difference between the measured output and the expected output via

$$\min_{X_{n+1}} \|Y_n - \hat{A}_n X_{n+1}\|_F^2. \quad (5.7)$$

For simplicity, we again consider this cost function at time  $n = 1$ , and assume  $\hat{A}_0 = A_0$ . Under these assumptions one finds,

$$\begin{aligned} \|Y_1 - \hat{A}_1 X_2\|_F^2 &= \|A_1 X_1 - (A_0 + dAX_1X_1^H)X_2\|_F^2 \\ &= \|(A_0 + dA)X_1 - (A_0 + dAX_1X_1^H)X_2\|_F^2 \\ &= \|A_0 X_1 + dAX_1 - (A_0 + dAX_1X_1^H)X_2\|_F^2 \\ \|Y_1 - \hat{A}_1 X_2\|_F^2 &= \|A_0(X_1 - X_2) + dA(X_1 - X_1X_1^H X_2)\|_F^2. \end{aligned}$$

With orthogonal inputs  $X_2 = X_1$ , this cost function will be *zero*. Similarly, if we assume instead no prior knowledge of the previous image, i.e.,  $\hat{A}_0 = \mathbf{0}$ , we find

$$\begin{aligned} \|Y_1 - \hat{A}_1 X_2\|_F^2 &= \|A_1 X_1 - (dAX_1X_1^H)X_2\|_F^2 = \|(A_0 + dA)X_1 - dAX_1X_1^H X_2\|_F^2 \\ &= \|A_0 X_1 + dA(X_1 - X_1X_1^H X_2)\|_F^2. \end{aligned}$$

Again, the cost function is minimized when  $X_2 = X_1$  and  $X_2$  is orthogonal. A third case is to assume the most recent estimate is a low-rank approximation of the first image,  $\hat{A}_0 = A_0 X_1 X_1^H$ .

In this case we find

$$\|Y_1 - \hat{A}_1 X_2\|_F^2 = \|A_1 X_1 - (A_0 X_1 X_1^H + dAX_1X_1^H)X_2\|_F^2$$

$$= \|A_1 X_1 - (A_1 X_1 X_1^H) X_2\|_F^2 = \|A_1 (X_1 - X_1 X_1^H X_2)\|_F^2.$$

Again, if  $X_2 = X_1$  and  $X_2$  is orthogonal the cost function is minimized. This shows that for a range of cases, from assuming complete knowledge, partial knowledge, and no knowledge of the previous image, this minimization problem choice results in another subspace trap.

In summary, most straight-forward approaches for selecting a new input vector set from previous output data and estimates results in the new inputs spanning the same subspace as the previous inputs. This bias towards previous inputs is due to the fact that changes in the estimate  $\hat{A}_n$  can only reflect the subspace supported by the output data,  $A_n X_n X_n^H$ . The key to escaping the trap is to choose new input vectors that “look in new places”. That is, for each new image acquisition,  $X_{n+1}$  should span some part of the previously untracked modes,  $(I - X_n X_n^H)$ . The following two sections show two ways to do this. One proposes and minimizes a cost function based on finding the right singular vectors of the true image  $A_n$  directly. The second introduces a temporal predictor to estimate  $A_{n+1}$  and uses the predicted image to determine appropriate acquisition vectors.

### 5.2.2 Escaping the subspace trap I: CG-St

The previous success of the SVD method [47], and the fact that the SVD is a limiting case of the ROI solution (§ 4.2.1), suggests that one may find useful input vectors by trying to identify the (orthogonal) right singular vectors of the image  $A_n$  directly. Horn and Johnson [18] show that solving  $\max_x \|Ax\|$  will identify the dominant right singular vector of  $A$ . Translating this problem to solve for multiple vectors  $x_k$  at once gives

$$\max_{X_{n+1}} \|A_n X_{n+1}\|_F^2 \equiv \min_{X_{n+1}} \|A_n (I - X_{n+1} X_{n+1}^H)\|_F^2. \quad (5.8)$$

The straight-forward approach is to substitute  $A_n$  with  $\hat{A}_n$ . However, as shown in Section 5.2.1, this tends to bias  $X_n$  towards the previous inputs as changes in the estimate,  $\hat{A}_n$ , can only reflect the subspace supported by the output data,  $A_n X_n X_n^H$ . To escape this subspace trap, we introduce a time lag in the right hand side of (5.8), substituting the *most recent* measurement

$Y_n = A_n X_n$  for the a-priori predicted output  $A_n X_{n+1}$ . We then substitute the image estimate  $\hat{A}_n$  for the first term in (5.8), which results in

$$\min_{X_{n+1}} \left\| \hat{A}_n - Y_n X_{n+1}^H \right\|_F^2. \quad (5.9)$$

To maintain the validity of the adaptive framework estimate update equation (5.5), the inputs  $X_{n+1}$  must remain orthogonal. Described below is a minimization algorithm suitable for such a task. The parameter space of all orthonormal matrices is known as the Stiefel manifold [11]. After a brief review of the conjugate gradient method, the method of constraining the conjugate gradient algorithm to the Stiefel manifold is presented.

### The conjugate gradient method

According to Golub and Van Loan [12], the method of conjugate gradients to solve linear systems was first described by Hestenes and Stiefel in 1952. It found widespread use as an iterative method in the 1970's due to its superior convergence properties. The derivation of the method described by Golub and Van Loan [12, Chpt. 10] is given below.

The derivation of the method is best described by considering the minimization of the function

$$\phi(x) = \frac{1}{2} x^T A x - x^T b \quad (5.10)$$

where  $x, b \in \mathbb{R}^n$  and  $A \in \mathbb{R}^{n \times n}$  is assumed to be positive definite and symmetric. Minimizing  $\phi(x)$  with respect to  $x$  gives  $x = A^{-1}b$ . Thus, solving the linear system  $Ax = b$  and minimizing  $\phi(x)$  are equivalent if  $A$  is symmetric and positive definite.

One of the simplest methods for numerically minimizing  $\phi(x)$  is the *method of steepest descent*. This method uses the negative gradient of the function at a current point  $x_n$  to determine a step direction, and then moves to a new point  $x_{n+1}$  that minimizes the function along the step direction. Analytically, the current solution estimate is updated via

$$x_{n+1} = x_n + \alpha r_n, \quad (5.11)$$

the search direction is determined via the *residual*

$$-\nabla\phi(x_n) = r_n = b - Ax_n, \quad (5.12)$$

and the step length is determined via

$$\operatorname{argmin}_{\alpha} \phi(x_n + \alpha r_n) = \operatorname{argmin}_{\alpha} [\phi(x_n) - \alpha r_n^T r_n + \frac{1}{2} \alpha^2 r_n^T A r_n]$$

with the solution

$$\alpha = (r_n^T r_n) / (r_n^T A r_n). \quad (5.13)$$

A summary of the steepest descent algorithm is given in the box below.

**Steepest Descent Algorithm**

To solve  $Ax = b$ , with  $b, x \in \mathbb{R}^n$  and where  $A \in \mathbb{R}^{n \times n}$  is assumed to be symmetric and positive definite

```

k=0; x = initial guess;
while  $r_k \neq 0$ ,
    k = k+1;
     $r_k = b - Ax_k$ ;           % compute residual
     $\alpha = (r_k^T r_k) / (r_k^T A r_k)$  % calculate step size
     $x_k = x_k + \alpha r_k$        % update solution
end;
```

One of the consequences of this technique is that consecutive residuals are orthogonal to each other.

$$\begin{aligned}
r_k^T r_{k-1} &= (b - Ax_k)^T r_{k-1} = (b - Ax_{k-1} - \alpha A r_{k-1})^T r_{k-1} \\
&= r_{k-1}^T r_{k-1} - \alpha r_{k-1}^T A^T r_{k-1} \\
&= r_{k-1}^T r_{k-1} - \frac{r_{k-1}^T r_{k-1}}{r_{k-1}^T A^T r_{k-1}} r_{k-1}^T A^T r_{k-1} \\
r_k^T r_{k-1} &= r_{k-1}^T r_{k-1} - r_{k-1}^T r_{k-1} = 0.
\end{aligned}$$

This occurs because the steepest descent algorithm actually oversteps the principal axes of the cost surface on each iteration. The result is that over the course of the algorithm the same search directions are used repeatedly with each new search direction  $r_k$  orthogonal to the last. This

behavior leads to a convergence time that can be prohibitively slow. A slow convergence time is especially apparent for systems with a large eigenvalue spread, i.e.,  $\lambda_{\max}/\lambda_{\min} \gg 1$  where  $\lambda_k$  are the eigenvalues of  $A$ .

Ideally, there is no requirement that the search directions be mutually orthogonal and this is the motivation for the conjugate gradient algorithm. The inefficient search behavior of the steepest descent method can be improved by changing the method by which search directions are chosen. In choosing these new search directions,  $p_k$ , one needs to ensure that  $p_k$  is *not* orthogonal to  $r_k$ , but rather the series  $\{p_1, p_2, \dots, p_k\}$  should span the same subspace as  $\{r_1, r_2, \dots, r_k\}$ .

Starting from the standard update equation,

$$x_k = x_{k-1} + \alpha_k p_k, \quad (5.14)$$

it follows that the parameter  $\alpha_k$  that minimizes  $\phi(x_{k-1} + \alpha_k p_k)$  is

$$\alpha_k = (p_k^T r_{k-1}) / (p_k^T A p_k). \quad (5.15)$$

The search direction itself is identified by considering the following: In order to construct an iterative algorithm, one would like to separate the minimization problem into two portions, one minimized previously up to point  $k-1$ , and the next point to find,  $k$ . The next solution point,  $x_k$ , can be represented with such a linear combination using the initial search point  $x_0$  and all subsequent search directions, i.e.,

$$x_k = x_0 + \sum_{n=1}^{k-1} \alpha_n p_n + \alpha_k p_k = x_0 + P_{k-1} y + \alpha_k p_k$$

where the columns of  $P_{k-1}$  are composed of the previous search directions,  $[p_1 \ p_2 \ \dots \ p_{k-1}]$ , and  $y$  contains the previous search step sizes,  $y^T = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_{k-1}]$ . Examining the function  $\phi(x_k)$  using this representation for  $x_k$  one finds

$$\begin{aligned} \phi(x_k) &= \frac{1}{2} (x_0 + P_{k-1} y + \alpha_k p_k)^T A (x_0 + P_{k-1} y + \alpha_k p_k) - (x_0 + P_{k-1} y + \alpha_k p_k)^T b \\ &= \frac{1}{2} (x_0 + P_{k-1} y)^T A (x_0 + P_{k-1} y) + \frac{1}{2} \alpha_k^2 p_k^T A p_k \end{aligned}$$

$$\begin{aligned}
& + \alpha_k p_k^T A(x_0 + P_{k-1}y) - (x_0 + P_{k-1}y)^T b - \alpha_k p_k^T b \\
\phi(x_k) & = \phi(x_0 + P_{k-1}y) + \frac{1}{2} \alpha_k^2 p_k^T A p_k + \alpha_k p_k^T A P_{k-1}y - \alpha_k p_k^T r_0.
\end{aligned}$$

If one requires that  $p_k^T A P_{k-1} = 0$ , then the cross-term,  $\alpha_k p_k^T A P_{k-1}y$ , in the above expression will cancel. This condition implies that each new search direction  $p_k$  be orthogonal to the subspace spanned by  $\{A p_0, A p_1, \dots, A p_{k-1}\}$ . With this condition, the minimization of  $\phi(x_k)$  can now be separated into two components, one dependent on the previously determined steps, and one dependent only on the new step.

$$\begin{aligned}
\min_{\alpha_k} \phi(x_k) & = \phi(x_0 + P_{k-1}y) + \min_{\alpha_k} \left[ \frac{1}{2} \alpha_k^2 p_k^T A p_k - \alpha_k p_k^T r_0 \right] \\
& \text{with } p_k \perp \{A p_0, A p_1, \dots, A p_{k-1}\}
\end{aligned}$$

The requirement

$$p_k^T A P_{k-1} = 0 \quad (5.16)$$

is known as the *conjugate gradient condition*. Motivated by the fact that the search directions  $\{p_1, \dots, p_k\}$  and the residuals  $\{r_1, \dots, r_k\}$  span the same subspace, and that the residuals can be determined in with an iterative update equation,  $r_k = r_{k-1} - \alpha_k A r_{k-1}$ , it follows that the new search directions can be determined in a similar fashion, i.e.,

$$p_k = r_{k-1} + \beta_k p_{k-1}. \quad (5.17)$$

Imposing the conjugate gradient condition,  $p_{k-1}^T A p_k = 0$ , on (5.17), and solving for  $\beta_k$  it follows that

$$\beta_k = -\frac{p_{k-1}^T A r_{k-1}}{p_{k-1}^T A p_{k-1}}. \quad (5.18)$$

This leads to the standard conjugate gradient method given in the box below.

A comparison of the two algorithms is given in Figure 5.1. For this example,

$$\begin{aligned}
A & = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix}^T = \begin{bmatrix} 1.5760 & 0.8653 \\ 0.8653 & 1.5406 \end{bmatrix}, \\
b & = u_1 \sigma_1 = \begin{bmatrix} 1.7312 \\ 1.6962 \end{bmatrix}, \quad \text{and} \quad x_0 = \begin{bmatrix} -1.3194 \\ 0.9312 \end{bmatrix}.
\end{aligned}$$

**Standard Conjugate Gradient**

```

 $x_0$  = initial guess
 $r_0 = b - Ax_0$ ,  $p_0 = 0$ , and  $k = 0$ 
while  $r_k \neq 0$ 
     $k = k + 1$ 
    if  $k = 1$ ,
         $\beta_k = 0$ ,
    else
         $\beta_k = -(p_{k-1}^T Ar_{k-1}) / (p_{k-1}^T Ap_{k-1})$ 
    end
     $p_k = r_{k-1} + \beta_k p_{k-1}$ 
     $\alpha_k = (p_k^T r_{k-1}) / (p_k^T Ap_k)$ 
     $x_k = x_{k-1} + \alpha_k p_k$ 
     $r_k = b - Ax_k$ 
end

```

The top figure shows the convergence behavior of the two algorithms. The bottom figure shows the trajectories of each algorithm over a contour plot of the function  $\phi(x)$ . For the  $A$  matrix specified,  $\lambda_{\max}/\lambda_{\min} = 3.4984$ , which implies an elliptical contour line as is seen in the contour plot of Figure 5.1. The lower figure also shows the orthogonality of successive search directions in the steepest descent method. The superior convergence behavior of the conjugate gradient method is clearly demonstrated in both figures.

The conjugate gradient algorithm can be made more computationally efficient by computing the residual recursively via  $r_k = r_{k-1} - \alpha_k Ap_k$  and substituting

$$r_{k-1}^T r_{k-1} = -\alpha_{k-1} r_{k-1}^T Ap_{k-1}$$

and

$$r_{k-2}^T r_{k-2} = \alpha_{k-1} p_{k-1}^T Ap_{k-1}$$

into the formula for  $\beta_k$ . This results in the conjugate gradient algorithm listed in Table 5.2, and is the foundation for the conjugate gradient on the Stiefel manifold (CG-St) algorithm developed by Edelman, Arás, and Smith.



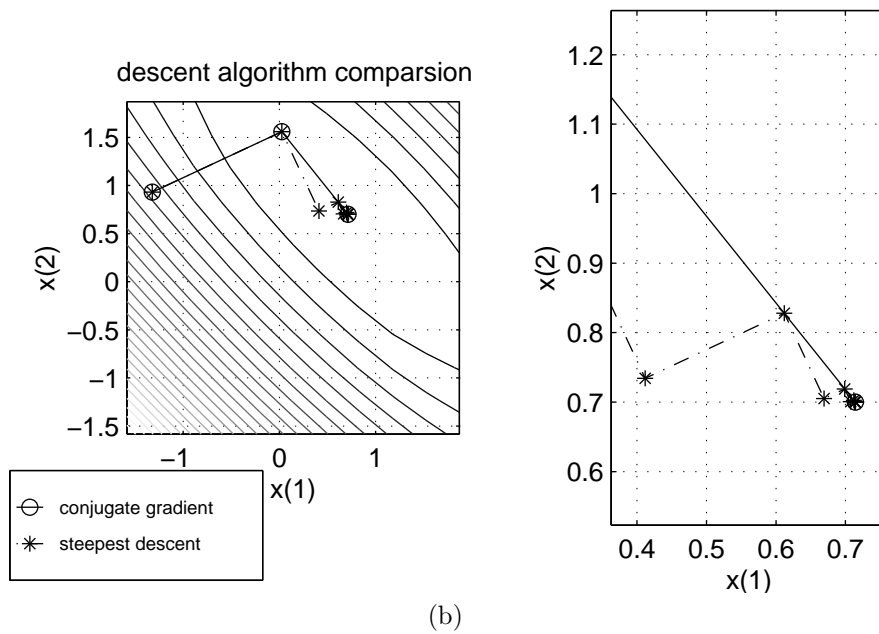
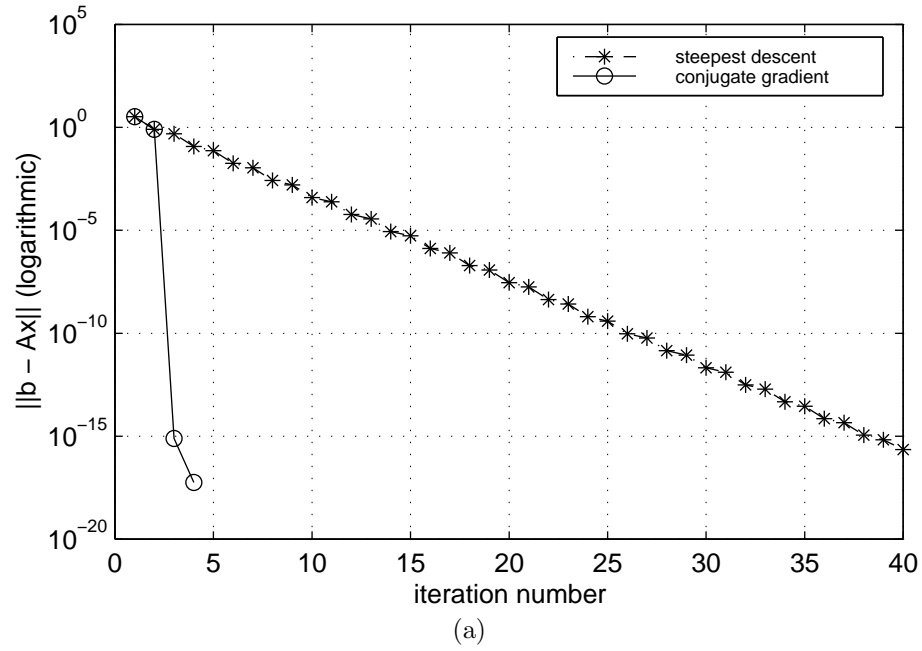


Figure 5.1: Comparison between the steepest descent and conjugate gradient methods. (a) Convergence Curve (b) Path through solution space

**Efficient Conjugate Gradient Algorithm**

Solves  $Ax = b$ , where  $x, b \in \mathbb{R}^n$  and  $A \in \mathbb{R}^{n \times n}$  is assumed to be positive definite and symmetric.

```

 $x_0$  = initial guess
 $r_0 = b - Ax_0$ ,  $p_0 = 0$ , and  $k = 0$ 
while  $r_k \neq 0$ 
     $k = k + 1$ 
    if  $k = 1$ ,
         $\beta_k = 0$ ,
    else
         $\beta_k = -(p_{k-1}^T Ar_{k-1}) / (p_{k-1}^T Ap_{k-1})$ 
    end
     $p_k = r_{k-1} + \beta_k p_{k-1}$ 
     $\alpha_k = (p_k^T r_{k-1}) / (p_k^T Ap_k)$ 
     $x_k = x_{k-1} + \alpha_k p_k$ 
     $r_k = b - Ax_k$ 
end

```

Table 5.2: Efficient conjugate gradient algorithm

There are many approaches for adapting non-symmetric systems to the conjugate gradient method. One possibility is to solve the system  $A^T Ax = A^T b$  instead of  $Ax = b$ . Another is to construct a coupled system

$$AA^T y = b \quad x = A^T y$$

and solve the “y space” problem using the conjugate gradient algorithm. A thorough review of these methods and others, including a discussion of numerical stability for each, is given in Chapter 10.4 of Golub and Van Loan [12].

**Conjugate gradient on the Stiefel manifold**

The conjugate gradient algorithm derived above is designed to solve a linear system of the form  $Ax = b$  with the unknown parameters given in a single vector  $x$ . For the purposes of dynamic MRI imaging, a “tall-skinny” matrix  $X$  of vectors would be more beneficial. A conjugate gradient algorithm containing this exact extension is described by Edelman, Ariás, and Smith (EAS) in

[11]. Specifically, they describe a conjugate gradient algorithm on the Stiefel manifold (CG-St) that is suitable for solving systems of the type  $AX = B$ , where  $A$  is symmetric and both  $X$  and  $B$  are tall-skinny matrices.

The Stiefel manifold,  $St(n, p)$ , is defined as the parameter space consisting of  $n$ -by- $p$  “tall-skinny” orthonormal matrices. For a point  $X$  on the manifold, this implies  $X^T X = I_p$ . The manifold may be embedded in the  $np$ -dimensional Euclidean space of all  $n$ -by- $p$  matrices. For a visual reference, when  $p = 1$ , the manifold is simply a sphere. When  $p = n$ , the manifold is the group of orthogonal matrices  $O_n$  that completely span the  $n$ -dimensional space.

The basic idea that leads to differential manifolds is that one would like “to select a family of a subcollection of neighborhoods so that a change of coordinates is always given by differentiable functions.” [3]. The beauty of operating on the Stiefel manifold is that the path given by any movement through Euclidean space will stay on the Stiefel manifold if the movement follows the set of differential equations describing the manifold. Drawing primarily from material presented by Edelman, Ariás, and Smith in [11], this section explores the equations describing the Stiefel manifold and describes the CG-St algorithm.

**TANGENT AND NORMAL SPACES:** An equation defining tangents to the Stiefel manifold at the point  $X$  is obtained by differentiating  $X^T X = I$ , which yields  $X^T \Delta + \Delta^T X = 0$ . Here  $\Delta$  refers to a matrix (subspace) tangent to the manifold. Note that  $X^T \Delta$  is skew-symmetric, i.e.,  $(X^T \Delta - \Delta^T X)/2 = X^T \Delta$ . This condition imposes  $p(p+1)/2$  constraints on  $\Delta$ . Thus, the vector space of all tangent vectors  $\Delta$  has dimension

$$np - \frac{p(p+1)}{2} = \frac{p(p-1)}{2} + p(n-p).$$

The normal space at a point  $X$  on a manifold is defined to be the set of all matrices that are orthogonal to the tangent space. If one views the Stiefel manifold as embedded in Euclidean space, one may choose the standard inner product

$$g_e(\Delta_1, \Delta_2) = \text{tr}\{\Delta_1^T \Delta_2\} = \sum_i (\Delta_1^T \Delta_2)_{ii} \quad (5.19)$$

in  $np$ -dimensional Euclidean space (hence the subscript  $e$ ) to test orthogonality. Thus, the normal space at a point  $X$  on the manifold consists of all points  $N$  which satisfy

$$\text{tr}\{\Delta^T N\} = 0$$

for all  $\Delta$  in the tangent space. It follows that the normal space is of dimension  $p(p+1)/2$ .

To implement the conjugate gradient condition (5.16) for the CG-St algorithm, the concept of subspace projection operators on the Stiefel manifold are defined here. If we denote  $\text{sym}(A) = (A + A^T)/2$  and  $\text{skew}(A) = (A - A^T)/2$ , then for a matrix  $Z$  of size  $n$ -by- $p$  we may write,

$$\pi_N(Z) = X \text{sym}(X^T Z) \quad (5.20)$$

as the projection of  $Z$  onto the normal space at the point  $X$  on the Stiefel manifold. Similarly,

$$\pi_T(Z) = X \text{skew}(X^T Z) + (I - XX^T)Z \quad (5.21)$$

as the projection of  $Z$  onto the tangent space at  $X$ . Equation (5.21) suggests that the matrices  $\Delta$  describing the tangential directions at  $X$  have the general form

$$\Delta = XA + X_\perp B = XA + (I - XX^T)C \quad (5.22)$$

where  $A$  is  $p$ -by- $p$  symmetric,  $B$  is  $(n-p)$ -by- $p$ ,  $C$  is  $n$ -by- $p$ ,  $B$  and  $C$  are both arbitrary, and  $X_\perp$  is any  $n$ -by- $(n-p)$  matrix such that  $XX^T + X_\perp X_\perp^T = I$ .

THE CANONICAL METRIC: The Euclidean metric given in Equation (5.19) is not necessarily suitable for use on the Stiefel manifold because it does not weigh each parameter of  $\Delta$  equally. This is seen by applying the metric to the general representation of the tangential vector given in (5.22)

$$g_e(\Delta, \Delta) = \text{tr}\{A^T A\} + \text{tr}\{BB^T\} = 2 \sum_{i < j} a_{ij}^2 + \sum_{ij} b_{ij}^2. \quad (5.23)$$

Because  $A$  is skew-symmetric, the Euclidean metric counts the independent degrees of freedom in  $A$  twice. EAS suggest that a more equitable metric would be  $g_c(\Delta, \Delta) = 1/2 \text{tr}\{A^T A\} + \text{tr}\{B^T B\}$ . Such a metric can be found by recognizing that at the point  $X = I_{n,p}$  (the first  $p$  columns of the

$n$ -by- $n$  identity matrix), the Euclidean metric can be modified as  $\text{tr}\{\Delta^T(I - 1/2I_{n,p}I_{n,p}^T)\Delta\}$  to give the equitable calculation above. Returning to the more general form gives

$$g_c(\Delta, \Delta) = \text{tr}\{\Delta^T(I - \frac{1}{2}XX^T)\Delta\} \quad (5.24)$$

which is called the *canonical metric* on the Stiefel manifold.

EMBEDDED GEODESICS: A geodesic is defined as the curve of shortest length between two points on a manifold. Typically, these curves are defined analytically through the use of an acceleration vector, and/or the calculus of variations. For example, stating that  $X(t)$  remains on the Stiefel manifold,  $St(n, p)$ , implies the condition  $X^T X = I_p \forall t$ . Taking two derivatives with respect to time, we find

$$X^T \ddot{X} + 2\dot{X}^T \dot{X} + \ddot{X}^T X = 0. \quad (5.25)$$

To be a geodesic,  $\ddot{X}(t)$  must be in the normal space at  $X(t)$ . That is, the acceleration vector must be normal to the geodesic surface. Pictorially, this is similar to a mass on a string being spun in a circle. The acceleration vector is constantly pointed to the center of the circle, and the path traced by the mass is a geodesic. From (5.20),  $XS$  is in the normal space if  $S$  is  $p$ -by- $p$  symmetric<sup>1</sup>. Thus, with  $\ddot{X} = XS$ , Equation (5.25) can be rewritten as

$$\ddot{X} + X(\dot{X}^T \dot{X}) = 0 \quad (5.26)$$

which is the differential equation that defines a geodesic on the Stiefel manifold.

In terms of the conjugate gradient algorithm, the geodesic description given in (5.26) is not very useful. EAS showed that it is possible to express the geodesic  $X(t)$  in terms of the current position  $X(t) = X$  and a direction  $\dot{X}(t) = P$ . Specifically, if  $X$  and  $P$  are  $n$ -by- $p$  matrices such that  $X^T X = I_p$  and  $A = X^T P$  is skew-symmetric, then the geodesic on the Stiefel manifold emanating from  $X$  in the direction  $P$  is given by the curve

$$X(t) = XM(t) + QN(t) \quad (5.27)$$

---

<sup>1</sup> $\pi_N(XS) = X\text{sym}(X^T XS) = X\text{sym}(S) = XS$  if  $S$  is symmetric

where

$$QR = (I - XX^T)P$$

is the compact QR decomposition of  $(I - XX^T)P$ , i.e.,  $Q$  is a tall-and-thin matrix.  $M(t)$  and  $N(t)$  are  $p$ -by- $p$  matrices given by the matrix exponential

$$\begin{pmatrix} M(t) \\ N(t) \end{pmatrix} = \exp \left\{ t \begin{pmatrix} A & -R^T \\ R & 0 \end{pmatrix} \right\} \begin{pmatrix} I_p \\ 0 \end{pmatrix}. \quad (5.28)$$

PARALLEL TRANSPORT EQUATION: The conjugate gradient algorithm makes explicit use of tangent and normal vectors through the course of finding a solution. Due to the curved nature of manifolds, vectors that were tangential at point  $X(0)$  may not be tangential when moved along a geodesic to the point  $X(t)$ . One can however, transport tangential vectors along a geodesic by removing the component in the normal space at each infinitesimal step. This process is known as *parallel transport*, and for the Stiefel manifold, EAS describe the translation of a tangential matrix  $P$  using the parameters of (5.27) and (5.28) via

$$\tau P(t) = PM(t) - X(t)R^T N(t)$$

Functionally,  $M(t)$  and  $N(t)$  play a role similar to the  $\cos(\theta)$  and  $\sin(\theta)$  components of a Givens rotation matrix and  $R$  acts as a scaling operator.

GRADIENT OF A FUNCTION: Finally, the conjugate gradient algorithm requires a computation of the gradient of the function. This computation is directly dependent on the choice of metric for the given parameter space. For a function  $F(X)$  on the Stiefel manifold  $St(n, p)$ , the gradient of  $F$  at  $X$  is defined to be the tangent vector  $\nabla F$  such that

$$g_c(\nabla F, \Delta) \equiv \text{tr}\{(\nabla F)^T (I - \frac{1}{2}XX^T)\Delta\} = \text{tr}\{F_X^T \Delta\} \quad (5.29)$$

for all tangent vectors  $\Delta$  at the point  $X$ , where  $F_X$  is the  $n$ -by- $p$  matrix of partial derivatives of  $F$  with respect to the elements of  $X$ , i.e.,

$$(F_X)_{ij} = \frac{\partial F}{\partial X_{ij}}. \quad (5.30)$$

To solve (5.29), one must find  $\nabla F$  such that the matrix lies in the normal space of the Stiefel manifold at point  $X$ , i.e.,  $X^T(\nabla F)$  is skew-symmetric. Solving (5.29) in this manner yields

$$\nabla F = F_X - XF_X^T X .$$

EAS note that this same expression may also be derived by differentiating  $F(X(t))$ , where  $X(t)$  is the geodesic equation given in (5.27).

THE CG-ST ALGORITHM: Taking each of the above equations that describe operations on the Stiefel manifold, one can now reconstruct the conjugate gradient algorithm on the Stiefel manifold. The mapping from the conjugate gradient algorithm of Table 5.2 to the CG-St algorithm of Table 5.3 is the following:

	CG	$\mapsto$	CG-St
The solution point	$x_k$	$\mapsto$	$X_k$
The search direction	$p_k$	$\mapsto$	$P_k$
The residual (gradient)	$r_k$	$\mapsto$	$G_k$
The inner product	$\langle u, v \rangle = u^T v$	becomes	$\text{tr}\{U^T(I - \frac{1}{2}XX^T)V\}$
The direction step	$x_k = x_{k-1} + \alpha_k p_k$	becomes	a minimization search along the geodesic described by $P_k$
The search update parameter	$\beta_k$	$\mapsto$	$\beta_k$

The CG-St algorithm is given in Table 5.3.

Finally, to implement the CG-St algorithm and numerically solve the dynamic MRI minimization problem given in Equation (5.9), we must determine the matrix of partial derivatives, (5.30), for the function  $F(X) = \|A - YX^H\|_F^2$ . Taking the partial derivative of  $F(X)$  with respect to  $X$  (see Appendix A.2) gives

$$F_X = -(A^T - XY^T)Y .$$

Examples using the CG-St algorithm for estimation dynamic sequences are given in Section 5.3.

**Conjugate Gradient for Minimizing  $F(X)$  on the Stiefel Manifold**

- Given  $X_0$  such that  $X_0^T X_0 = I$ , compute  $G_0 = F_{X_0} - X_0 F_{X_0}^T X_0$  and set  $P_0 = -G_0$ .
- For  $k = 0, 1, \dots$ ,
  - Compute the compact QR decomposition  $QR = (I - X_k X_k^T) P_k$ .
  - Minimize  $F(X_k(t))$  over  $t$  where

$$X_k(t) = X_k M(t) + Q N(t),$$

$A = X_k^T P_k$ ,  $M(t)$  and  $N(t)$  are  $p$ -by- $p$  matrices given by the  $2p$ -by- $2p$  matrix exponential given in Equation (5.28).

- Set  $t_k = t_{\min}$  and  $X_{k+1} = X_k(t_k)$ .
- Compute  $G_{k+1} = F_{X_{k+1}} - X_{k+1} F_{X_{k+1}}^T X_{k+1}$ .
- Parallel transport the tangent vector  $P_k$  to the point  $X_{k+1}$ :

$$\tau P_k = P_k M(t_k) - X_k R^T N(t_k). \quad (5.31)$$

Set  $\tau G_k := G_k$  or 0, which is *not* parallel.

- Compute the new search direction

$$P_{k+1} = -G_{k+1} + \beta_k \tau P_k \quad \text{where} \quad \beta_k = \frac{\langle G_{k+1} - \tau G_k, G_{k+1} \rangle}{\langle G_k, G_k \rangle}$$

and  $\langle \Delta_1, \Delta_2 \rangle = \text{tr}\{\Delta_1^T (I - \frac{1}{2} X X^T) \Delta_2\}$ .

- Reset  $P_{k+1} = -G_{k+1}$  if  $k+1 \equiv 0 \pmod{p(n-p) + p(p-1)/2}$ .

Table 5.3: Conjugate gradient for minimizing  $F(X)$  on the Stiefel manifold

### 5.2.3 Escaping the subspace trap II: Image prediction

Section 5.2.1 described the subspace trap, and showed that methods to find new input vectors from the estimates of prior images typically fails to provide useful new inputs. A close examination of the image reconstruction methods in Table 5.1 of Section 5.1 shows that it may be highly beneficial to guess at the composition of the next image and choose new input vectors based on that guess. This implies using an image predictor in the image estimation framework, and is the subject of the current section.

Specifically, if one constructs a predicted version,  $\tilde{A}_{n+1}$ , of the next true image in the sequence,  $A_{n+1}$ , the right singular vectors of the difference matrices in Table 5.1 should provide useful inputs



$X_{n+1}$  to acquire and reconstruct an estimate,  $\hat{A}_{n+1}$  of the next true image in the sequence. The design choices for this method concern the following: the prediction mechanism itself, the inputs into the predictor, and how to choose new vectors based on the predicted image. Each of these design parameters are explored in more detail below.

### Prediction methods

Because we are interested in predicting images, the prediction methods falls into two categories: temporal and spatial. In both cases, the prediction method is best guided by a model of the image dynamics. Examples include modeling the changes as an optical flow, identifying image sub-blocks and estimating the motion of each, or simply applying a linear model to the pixel changes and producing a temporal extrapolation using a linear fit to previous pixel estimate values. A more complete description of potential image prediction methods is given below in Chapter 6. For now, we will focus primarily on pixel-by-pixel extrapolation along the temporal dimension.

Let  $\tilde{a}_{ij,n+1}$  represent the pixel at row  $i$  and column  $j$  for the predicted image  $\tilde{A}_{n+1}$ . The pixel can be represented as a linear combination of previous pixels via

$$\tilde{a}_{ij,n+1} = \sum_{k=0}^{(t-1)} c_k \check{a}_{ij,n-k}. \quad (5.32)$$

where  $\check{a}_{ij,n}$  is the pixel at position  $(i, j)$  for the image matrix  $\check{A}_n$  which is used as input to the predictor. This model could be extended as well to include weighted contributions from neighboring pixels via

$$\tilde{a}_{ij,n+1} = \sum_{p=i-u}^{i+u} \sum_{q=j-v}^{j+v} \sum_{k=0}^{t-1} c_{pq,k} \check{a}_{pq,n-k}$$

Consider for now a limited case of (5.32). If we assume that the image sequence is uniformly sampled in time, the extrapolation performed by the linear predictor can be written in closed form. This allows the predetermination of the prediction coefficients  $c_k$  used to generate the predicted image.

The general format for pixel-by-pixel extrapolation using a linear model along the temporal

domain is described via  $y(t) = mt + b = [t \ 1] [m \ b]^T$ . One can generalize this expression for any number of points,  $t$ , for which the line will be fit. Writing this equation for the each point in the set we arrive at the system of equations for a least squares fit.

$$\begin{bmatrix} \check{a}_{ij,n} \\ \check{a}_{ij,n-1} \\ \vdots \\ \check{a}_{ij,n-t+1} \end{bmatrix} = \begin{bmatrix} t-1 & 1 \\ t-2 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = Z \begin{bmatrix} u \\ v \end{bmatrix}$$

For each pixel position,  $(i, j)$ , this same system holds. Thus, the expression for extrapolating the entire image is

$$\begin{bmatrix} \text{vec}\{\check{A}_n\}^T \\ \text{vec}\{\check{A}_{n-1}\}^T \\ \vdots \\ \text{vec}\{\check{A}_{n-t+1}\}^T \end{bmatrix} = Z \begin{bmatrix} u_{0,0} & u_{1,0} & \cdots & u_{m-1,n} & u_{m,n} \\ v_{0,0} & v_{1,0} & \cdots & v_{m-1,n} & v_{m,n} \end{bmatrix} = ZW$$

We choose to solve this system using the pseudo-inverse of  $Z^\dagger = (Z^T Z)^{-1} Z^T$ . Since  $Z$  only depends on the number of images  $t$  used for the prediction, this can be determined ahead of time.

Furthermore, each pixel in the predicted image is determined via

$$\text{vec}\{\tilde{A}_{n+1}\}^T = [t \ 1] W = [t \ 1] (Z^T Z)^{-1} Z^T \begin{bmatrix} \text{vec}\{\check{A}_n\}^T \\ \text{vec}\{\check{A}_{n-1}\}^T \\ \vdots \\ \text{vec}\{\check{A}_{n-t+1}\}^T \end{bmatrix}.$$

The expression  $[t \ 1] (Z^T Z)^{-1} Z^T$  depends only on  $t$ . Thus, if the image sequence is uniformly sampled in time, the equation describing the image prediction from a linear combination of prior images can be determined a-priori. Given  $\sum_{i=0}^{t-1} i^2 = t(t-1)(2t-1)/6$ ,  $\sum_{i=0}^{t-1} i = t(t-1)/2$ , and denoting  $\check{A}$  as the predictor input, the predictor output in closed form for a given value of  $t$  is

$$\tilde{A}_{n+1} \Big|_t = \sum_{k=0}^{t-1} \frac{12}{t^2(t^2-1)} \left[ \frac{t(t+1)}{2} k - \frac{t(t-1)(t+1)}{6} \right] \check{A}_{n-(k-(t-1))} = \sum_{j=0}^{t-1} \left[ \frac{4}{t} - j \frac{6}{t(t-1)} \right] \check{A}_{n-j}.$$

A few simple cases are given in Table 5.4 below.

$$\begin{aligned}
2 \text{ points: } & \tilde{A}_{n+1} = 2\check{A}_n - \check{A}_{n-1} \\
3 \text{ points: } & \tilde{A}_{n+1} = \frac{4}{3}\check{A}_n + \frac{1}{3}\check{A}_{n-1} - \frac{2}{3}\check{A}_{n-2} \\
4 \text{ points: } & \tilde{A}_{n+1} = \check{A}_n + \frac{1}{2}\check{A}_{n-1} - \frac{1}{2}\check{A}_{n-3} \\
5 \text{ points: } & \tilde{A}_{n+1} = \frac{4}{5}\check{A}_n + \frac{1}{2}\check{A}_{n-1} + \frac{1}{5}\check{A}_{n-2} - \frac{1}{10}\check{A}_{n-3} - \frac{2}{5}\check{A}_{n-4}
\end{aligned}$$

Table 5.4: Predetermined equations for image prediction from uniformly sampled image estimates

### Predictor input

In general, any reasonable representation of the image at time  $n$  can be used as input into the predictor. This includes the full keyhole-style image estimate,  $\hat{A}_n = Y_n X_n^H + \hat{A}_{n-1}(I - X_n X_n^H)$ , the instantaneous image reconstruction  $Y X^H$ , a low-rank approximation of the image, or a thresholded version of any of the above.

The examples shown in Section 5.3 compare only the keyhole-style image estimate  $\hat{A}_n = Y_n X_n^H + \hat{A}_{n-1}(I - X_n X_n^H)$ , and the instantaneous image reconstruction  $Y X^H$  as predictor inputs. As shown experimentally in Section 5.3, the consistently best results are obtained by using the instantaneous image reconstruction as the predictor input. An analysis of why this is so is provided below.

### New input vector identification

Barring omniscient knowledge of the next true image in the sequence, as is suggested by the theoretically optimal choices in Table 5.1, the key to successful selection of new input vectors is to consistently probe the “correct” system subspace without perpetually probing the “same” system subspace. The “correct” inputs truly depend on the image reconstruction method used. For the low-rank method of (5.3), the best inputs would be the right singular vectors of the next true image  $A_{n+1}$ . For the keyhole method of (5.4), the best input vectors are the right singular vectors of the difference matrix  $(A_{n+1} - A_0)$ . Similarly, for the general adaptive framework given in (5.5),

the “best” vectors are the right singular vectors of  $(A_{n+1} - \hat{A}_{n-1})$ .

In each of these cases, a more practical approach is to replace the true image  $A_{n+1}$  with a predicted image  $\tilde{A}_{n+1}$ , before proceeding to find the new input vectors. However, care must be taken to stay out of the subspace trap, even when using a predicted image to chose the new vectors. For example, consider the case of predicting the next image from two previous image estimates. With the inputs to the predictor denoted  $\check{A}_n$ , Table 5.4 specifies

$$\tilde{A}_{n+1} = 2\check{A}_n - \check{A}_{n-1}.$$

For the case when the predictor input is the adaptive framework estimate,  $\hat{A}_n = Y_n X_n^H + \hat{A}_{n-1}(I - X_n X_n^H)$ , we find

$$\begin{aligned} (\tilde{A}_{n+1} - \hat{A}_n) &= (2\hat{A}_n - \hat{A}_{n-1}) - \hat{A}_n = \hat{A}_n - \hat{A}_{n-1} \\ &= A_n X_n X_n^H + \hat{A}_{n-1}(I - X_n X_n^H) - \hat{A}_{n-1} \\ (\tilde{A}_{n+1} - \hat{A}_n) &= A_n X_n X_n^H - \hat{A}_{n-1} X_n X_n^H = (A_n - \hat{A}_{n-1}) X_n X_n^H \end{aligned} \quad (5.33)$$

The expression  $X_n X_n^H$  on the right side of the matrix product expression in (5.33) is a subspace projection becuase  $X_n$  is orthogonal. Thus, the right singular vectors of  $(\tilde{A}_{n+1} - \hat{A}_n)$  will span the same subspace as  $X_n$ . This implies that choosing new input vectors from the SVD of  $(\tilde{A}_{n+1} - \hat{A}_n)$  will give vectors spanning the same subspace at every time  $n$  — a subspace trap.

Considering the same case, but using the instantaneous reconstruction,  $\hat{A}_n = Y_n X_n^H$ , for the predictor input we find

$$\begin{aligned} (\tilde{A}_{n+1} - \hat{A}_n) &= (2\hat{A}_n - \hat{A}_{n-1}) - \hat{A}_n \\ &= 2Y_n X_n^H - Y_{n-1} X_{n-1}^H - \hat{A}_n \\ &= 2A_n X_n X_n^H - A_{n-1} X_{n-1} X_{n-1}^H - (Y_n X_n^H - \hat{A}_{n-1}(I - X_n X_n^H)) \\ (\tilde{A}_{n+1} - \hat{A}_n) &= A_n X_n X_n^H - A_{n-1} X_{n-1} X_{n-1}^H + \hat{A}_{n-1}(I - X_n X_n^H). \end{aligned}$$

This shows that the right singular vectors of  $(\tilde{A}_{n+1} - \hat{A}_n)$  will not necessarily be biased towards  $X_n$ . Depending on the similarities between the dominant subspaces of  $A_n$ ,  $A_{n-1}$ , and  $\hat{A}_{n-1}$ , one

can expect that with  $Y_n X_n^H$  as the predictor input, and selecting new input vectors from the right singular vectors of  $(\tilde{A}_{n+1} - \hat{A}_n)$ , one will find new input vectors that lead to a high quality estimate of the image sequence without falling into the subspace trap. This conjecture is confirmed in the simulation experiments shown in Section 5.3.

### 5.3 Method comparison examples

To properly show the utility of the two new methods given in Sections 5.2.2 and 5.2.3, this section shows a comparison between the CG-St and Linear Prediction methods and those methods proposed previous by other dynamic MRI researchers, specifically the Fourier Keyhole (FK) method, the keyhole SVD method, and the RIGR method. A brief discussion on the implementation of each of these methods is given below. Finally, these results are compared to the theoretically optimal solution using the adaptive framework, as given in Table 5.1 of Section 5.1.

In the Fourier Keyhole (FK) method, the full  $k$ -space data set,  $A_{ref}$ , of the first image in the sequence is acquired. Each additional image is reconstructed by first acquiring a limited number lines of  $k$ -space, typically associated with the lowest frequency components. An estimate of the  $k$ -space data matrix for the new image is constructed by combining the newly acquired data for the low-frequency components with data from the reference image for the high-frequency components. Analytically, this can be described via the linear system model (3.10) of Section 3.2 as

$$\hat{A}_n = Y_n X X^H + A_{ref}(I - X X^H) . \quad (5.34)$$

The RIGR method uses the same Fourier basis vectors for the data acquisition as the FK method. The estimate is formed by applying a spatial profile envelope from the reference image to a linear combination of the output data. A complete description of the method is given in Section 3.1.2.

The keyhole SVD method (SVD) used for comparison in this section uses the same estimate update equation as FK, (5.34). However, in this case, the input vectors  $X$  are chosen from the

right singular vectors of  $A_{ref}$  rather than the lowest frequency Fourier basis vectors.

The CG-St method uses the adaptive framework update equation

$$\hat{A}_n = Y_n X_n X_n^H + \hat{A}_{n-1} (I - X_n X_n^H). \quad (5.35)$$

The input vectors used to acquire each new image in the sequence are identified by the minimization of

$$X_{n+1} = \underset{X}{\operatorname{argmin}} \|\hat{A}_n - Y_n X^H\|_F^2$$

using the conjugate gradient descent algorithm on the Stiefel manifold, as described by Edelman, Ariás, and Smith in [11]. A complete description of the CG-St method is given in Section 5.2.2.

The linear prediction ( $\operatorname{lp}(\cdot)$ ) methods use the adaptive framework update equation (5.35) as well. For this method, the input vectors are identified by first determining a predicted estimate of the next image in the sequence  $\tilde{A}_{n+1}$ , and then identifying the matrix  $X$  using the dominant right singular vectors of the difference matrix  $(\tilde{A}_{n+1} - \hat{A}_n)$ . For the results shown in these examples, the predicted image is constructed from a linear combination of three past image estimates  $\check{A}_k$  via

$$\tilde{A}_{n+1} = \frac{4}{3}\check{A}_n + \frac{1}{3}\check{A}_{n-1} - \frac{2}{3}\check{A}_{n-2}$$

A complete review of the linear predictive method is given in Section 5.2.3. For this comparison, two different inputs are used in the linear predictor. One is the keyhole-style estimate  $\check{A}_n = Y_n X_n^H + \hat{A}_n (I - X_n X_n^H)$ , denoted  $\operatorname{lp}(\text{Aest})$ . The second is the instantaneous estimate  $\check{A}_n = Y_n X_n^H$ , denoted  $\operatorname{lp}(YX^H)$ .

Each of these methods are compared for a variety of dynamic sequences. These sequences include a contrast change sequence showing a contrast agent being absorbed by a tumor, and a motion sequence showing a “needle” being inserted into a grapefruit. For each sequence type, both synthetic sequences and a sequence constructed from real MRI data are compared. The synthetic sequences provide complete control over the features of the sequence, including limiting the amount of jitter and full image intensity and contrast change between frames. This allows one to identify

those algorithms that are best suited for a particular change. The real data sequences are provided to test the clinical utility of the methods.

### Contrast change

The first dynamic sequence used for comparison shows a contrast change. The data were originally acquired from a clinical tumor identification experiment showing the flow of a contrast agent. In the sequence, the tumor is visible inside the knee of the patient, just behind the kneecap. The tumor is clearly visible in the fortieth frame of the sequence, which is shown in Figure 5.2. This figure also illustrates the  $72 \times 72$  pixel region of interest used in the method comparison below. The ROI is shown with a standard intensity map. The region outside of the ROI is shown using an inverse intensity map. The original data contained real valued intensity values for each of the 60 images in the sequence.

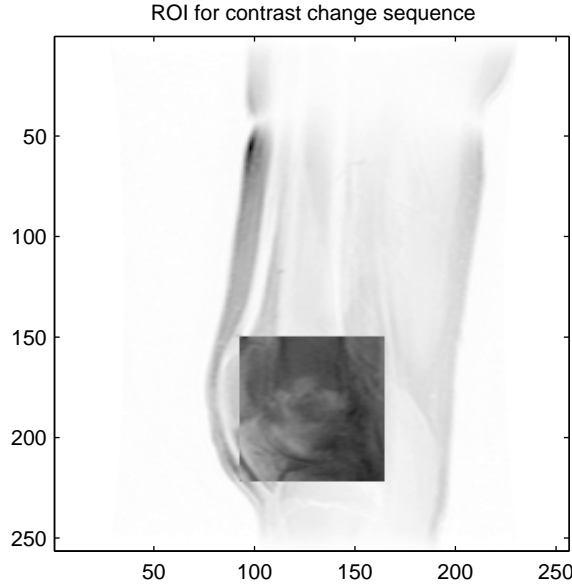


Figure 5.2: Contrast change example reference image showing ROI

To simulate the sequence, two frames (Image 15 and Image 34) were chosen from the original data set. The difference matrix between the two was calculated, and a fractional portion of this

difference matrix was added to the reference image, Image 15, to form a twenty frame image sequence. The fractional addition followed an arctangent curve to closely simulate the smooth transition from the “no contrast agent” state to the “full contrast agent” state shown in the original image sequence.

In Matlab, the commands were:

```
>> d = Image34 - Image15;
>> t = atan( 0.5*([1:20]-10));
>> ImgSeq = Image15(:)*ones(size(t)) + d(:)* 0.5*( 1 - t./min(t) );
```

The number of inputs  $r$  used to acquire the sequence was determined from a partial sum of the singular values. Specifically, the value of  $r$  was chosen such that

$$\frac{\sum_{n=1}^r \sigma_n}{\sum_{n=1}^N \sigma_n} = 85\% \quad (5.36)$$

where  $\sigma_n$  are the singular values of the  $72 \times 72$  square region of interest in the first image in the sequence. With this criterion,  $r = 8$  was selected for the knee-tumor contrast sequences. The choice implies that if this sequence were to be acquired using a MRI scanner, one could expect the acquisition time to be  $8/72 = 11.1\%$  the total acquisition time of the original sequence.

Figure 5.3 shows the original synthetic sequence. The time progression is shown first left-to-right, and then top-to-bottom. Thus, the twenty images in the sequence appear in the following order:

$A_0$	$A_1$	$A_2$	$A_3$	$A_4$
$A_5$	$A_6$	$A_7$	$A_8$	$A_9$
$A_{10}$	$A_{11}$	$A_{12}$	$A_{13}$	$A_{14}$
$A_{15}$	$A_{16}$	$A_{17}$	$A_{18}$	$A_{19}$

Figure 5.4 shows the relative error for each method over the course of the synthetic sequence. For this example, the Fourier based methods, FK and RIGR, outperform the linear system based methods. In fact, the performance of the Fourier based methods are remarkably similar. This is



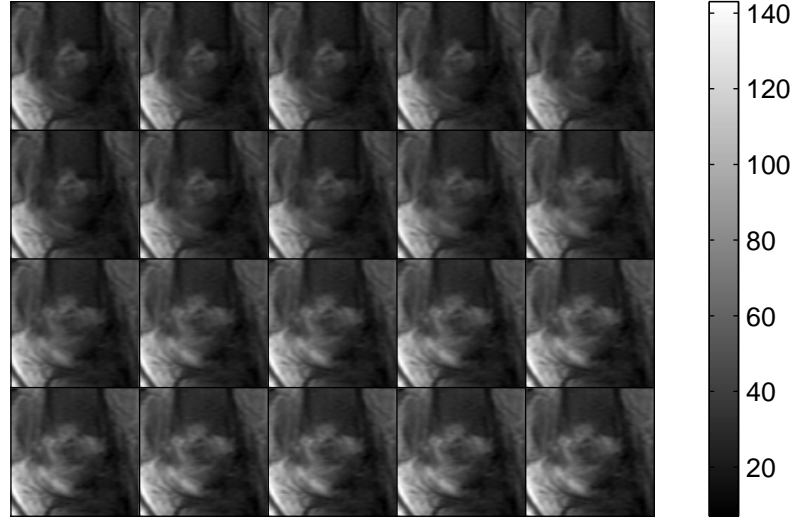


Figure 5.3: Original synthetic contrast change sequence

confirmed by visual inspection of the estimate sequences shown in Figures 5.6 through Figure 5.11. The time placement of each image for these figures occurs in the same order as the original image sequence. The reference image in each case is shown darker to differentiate it from the estimate images, and appears at time  $n = 0$  in the upper left corner of the image. Both the estimate  $\hat{A}_n$  and the absolute estimate error  $E_n = |A_n - \hat{A}_n|$  are shown.

(a)	$A_{ref}$	$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{A}_4$
	$\hat{A}_5$	$\hat{A}_6$	$\hat{A}_7$	$\hat{A}_8$	$\hat{A}_9$
	$\hat{A}_{10}$	$\hat{A}_{11}$	$\hat{A}_{12}$	$\hat{A}_{13}$	$\hat{A}_{14}$
	$\hat{A}_{15}$	$\hat{A}_{16}$	$\hat{A}_{17}$	$\hat{A}_{18}$	$\hat{A}_{19}$

and (b)

<b>0</b>	$ E_1 $	$ E_2 $	$ E_3 $	$ E_4 $
$ E_5 $	$ E_6 $	$ E_7 $	$ E_8 $	$ E_9 $
$ E_{10} $	$ E_{11} $	$ E_{12} $	$ E_{13} $	$ E_{14} $
$ E_{15} $	$ E_{16} $	$ E_{17} $	$ E_{18} $	$ E_{19} $

Note that each image set is shown using the full dynamic range of the color map to show the as much image detail as possible. This is especially apparent when comparing the absolute error image sequences. For example, the maximum estimate error is around 5.5 for the theoretically optimal acquisition where as the maximum keyhole SVD acquisition error is close to 40. If these

two images were shown using the same color map, the optimal error image would be very light and show very little detail.

Of particular interest is the theoretically optimal estimate sequence shown in Figure 5.5. The estimate error images show few features, appearing almost spatially white noise. The magnitude of the estimate error is also smaller than the realizable methods by nearly a factor of ten. This indicates that low-order acquisition methods have tremendous potential in reducing the acquisition time of dynamic MRI sequences.

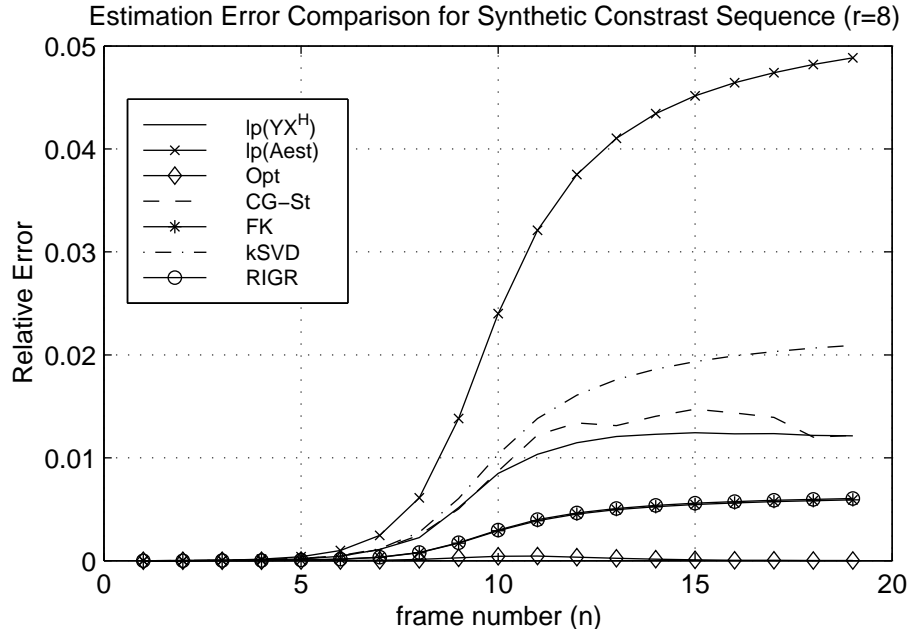
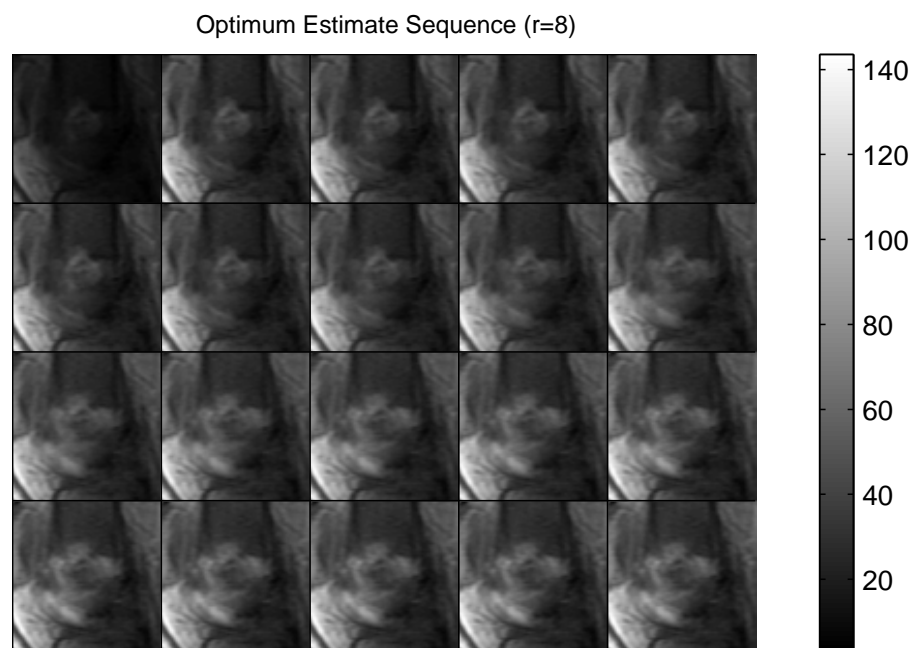
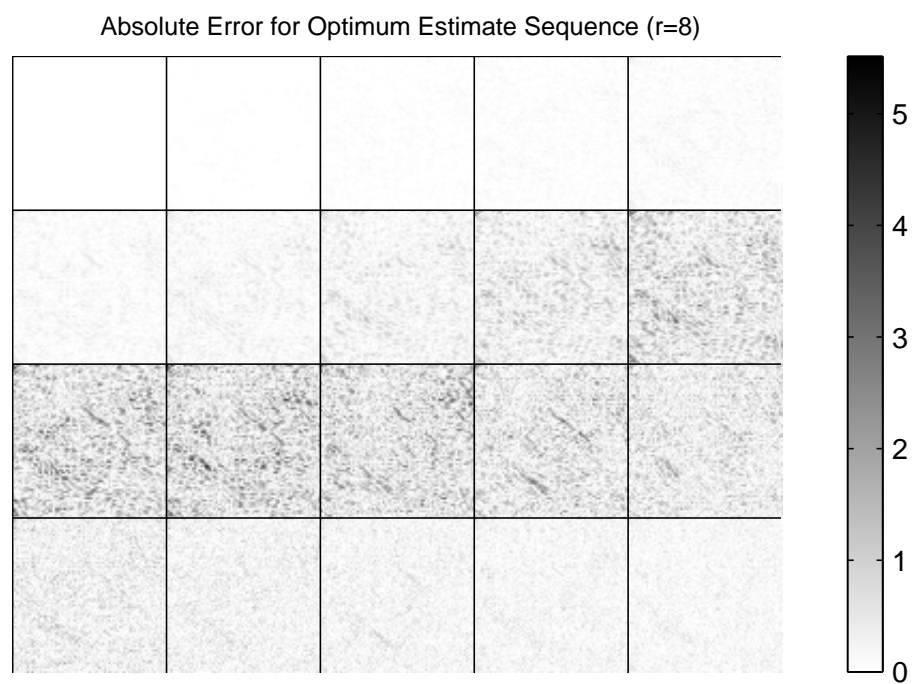


Figure 5.4: Low-order acquisition method comparison showing the relative error for the synthetic contrast change sequence

The results for the synthetic sequence shown previously are consistent with acquisition simulations using the original knee-tumor contrast sequence image data. A comparison of the relative error is given in Figure 5.12. This shows that again the Fourier based methods outperform the other realizable methods. Two of the new methods proposed in this thesis, CG-St and  $\text{lp}(YX^H)$ , perform better than the SVD method. The third,  $\text{lp}(\text{Aest})$ , does not due to the bias of the input selection towards previous inputs as discussed in Section 5.2.1.

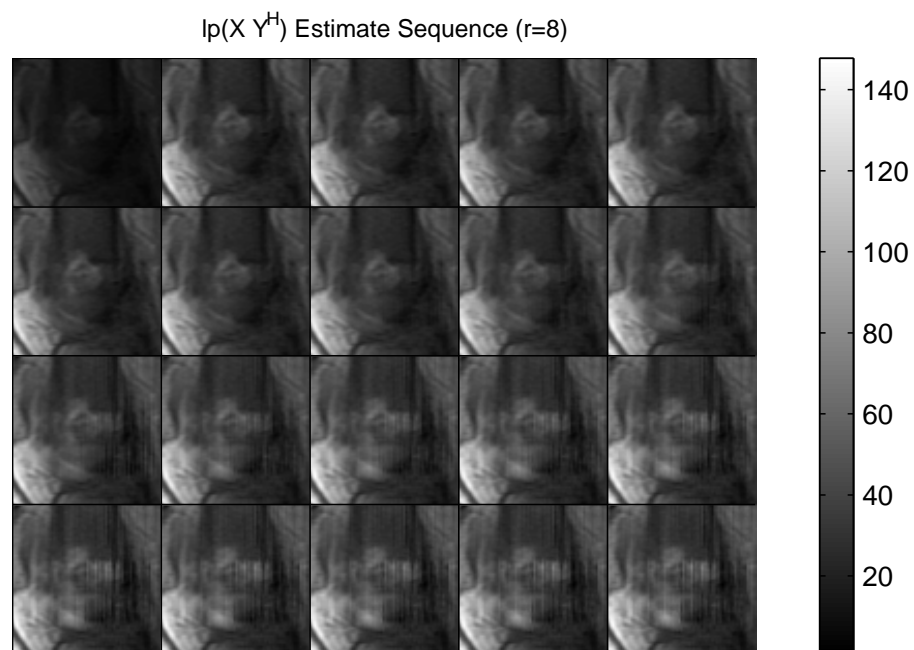


(a)

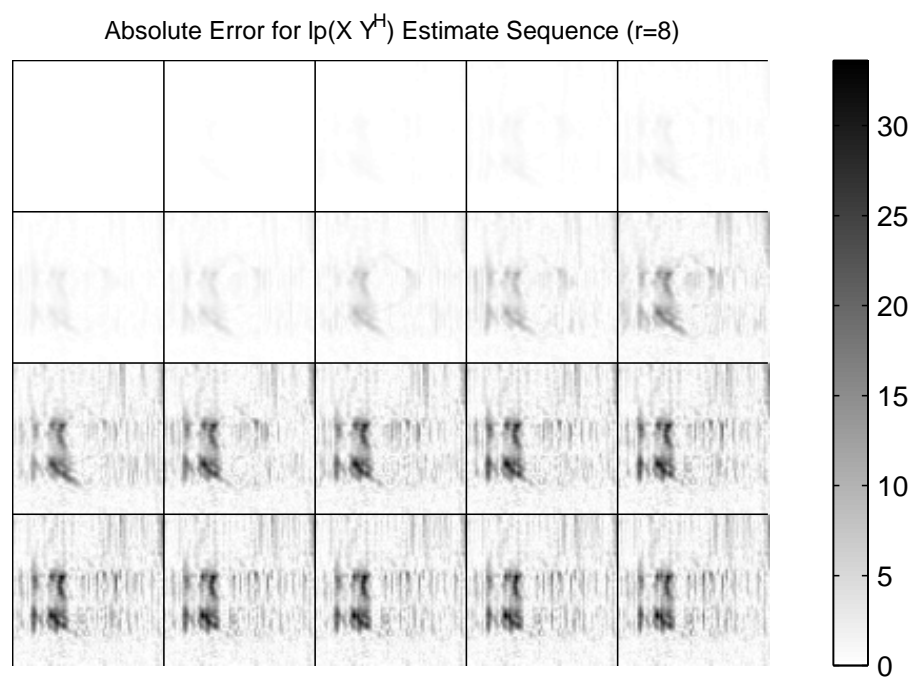


(b)

Figure 5.5: Simulated synthetic contrast change sequence acquisition using the Optimal method:  
 (a) Estimated Images and (b) Absolute Estimate Error

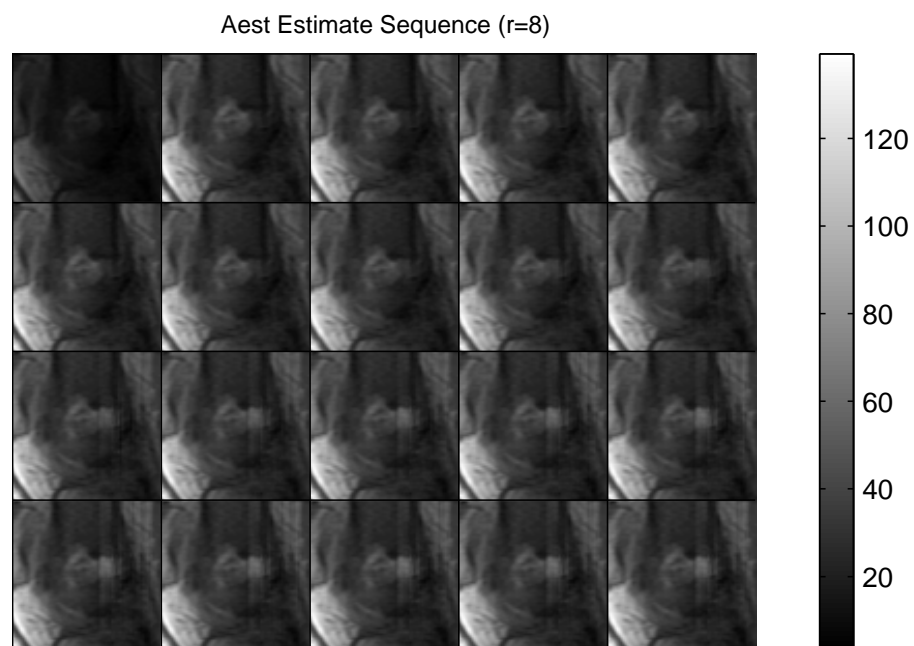


(a)

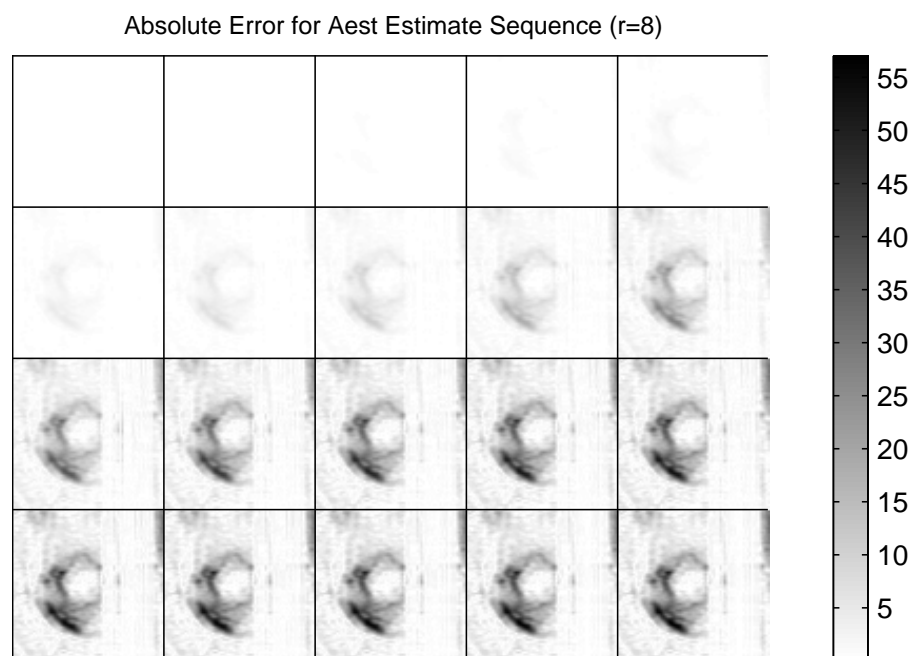


(b)

Figure 5.6: Simulated synthetic contrast change sequence acquisition using the Linear Predictor  $[\text{lp}(Y X^H)]$  Method: (a) Estimated Images and (b) Absolute Estimate Error

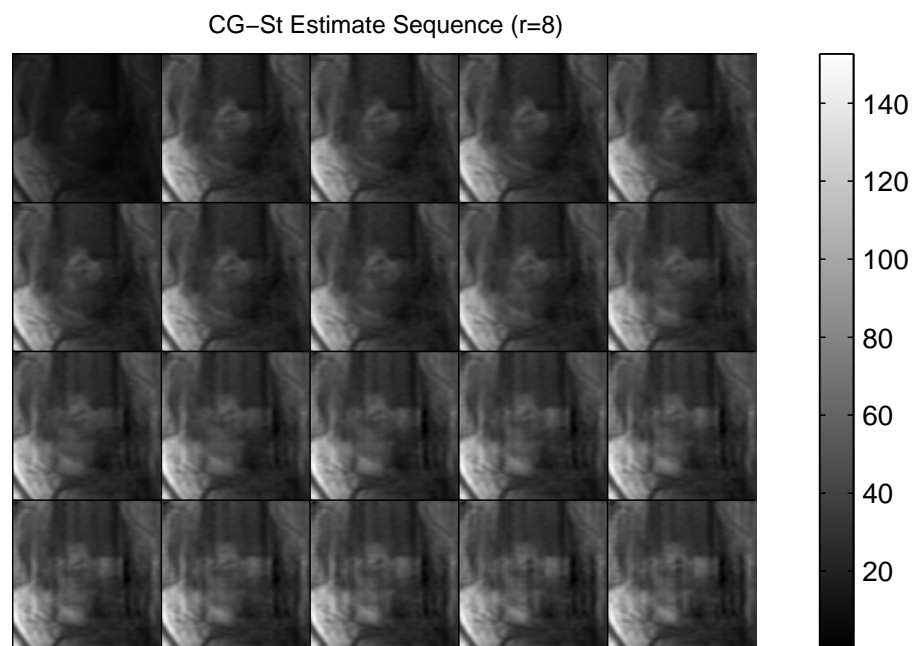


(a)

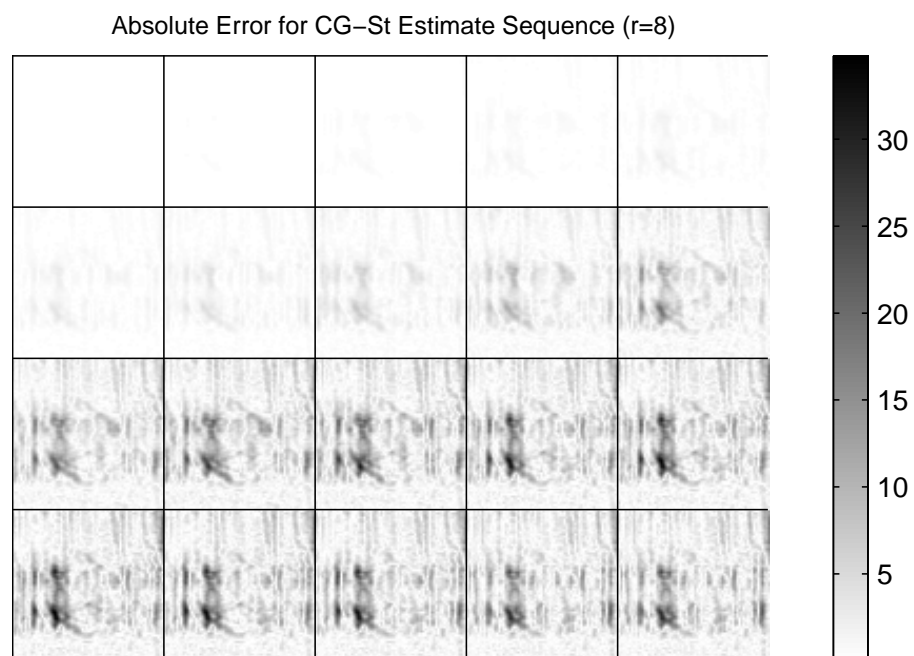


(b)

Figure 5.7: Simulated synthetic contrast change sequence acquisition using the Linear Predictor [lp(Aest)] Method: (a) Estimated Images and (b) Absolute Estimate Error

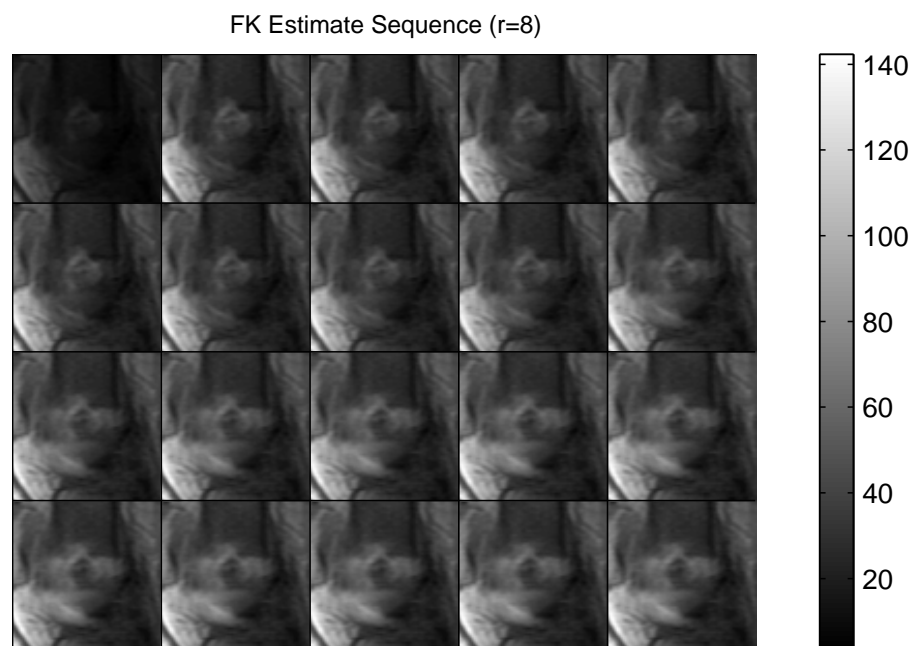


(a)

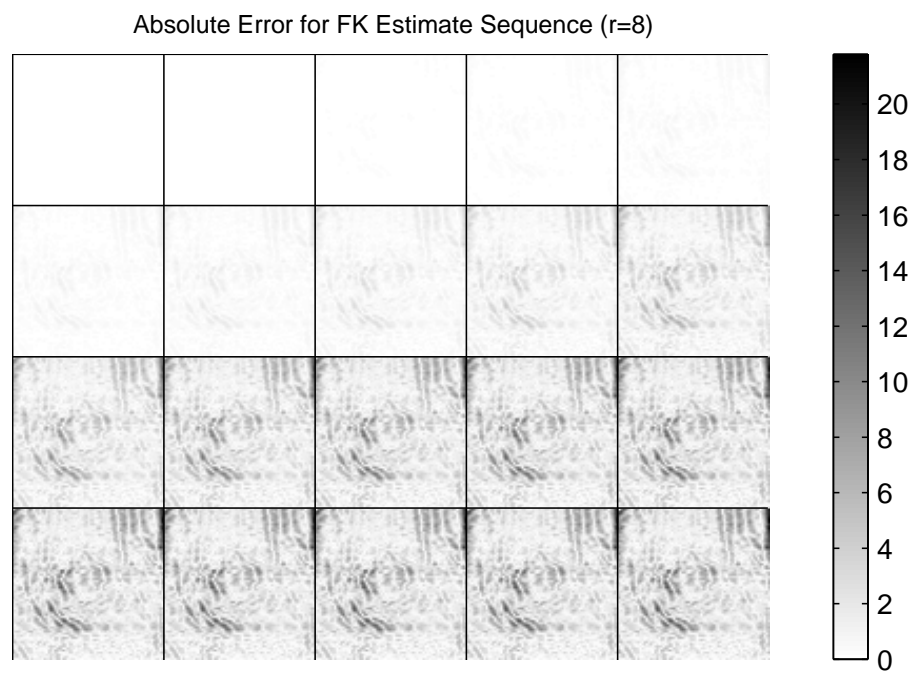


(b)

Figure 5.8: Simulated synthetic contrast change sequence acquisition using the CG-St method: (a) Estimated Images and (b) Absolute Estimate Error

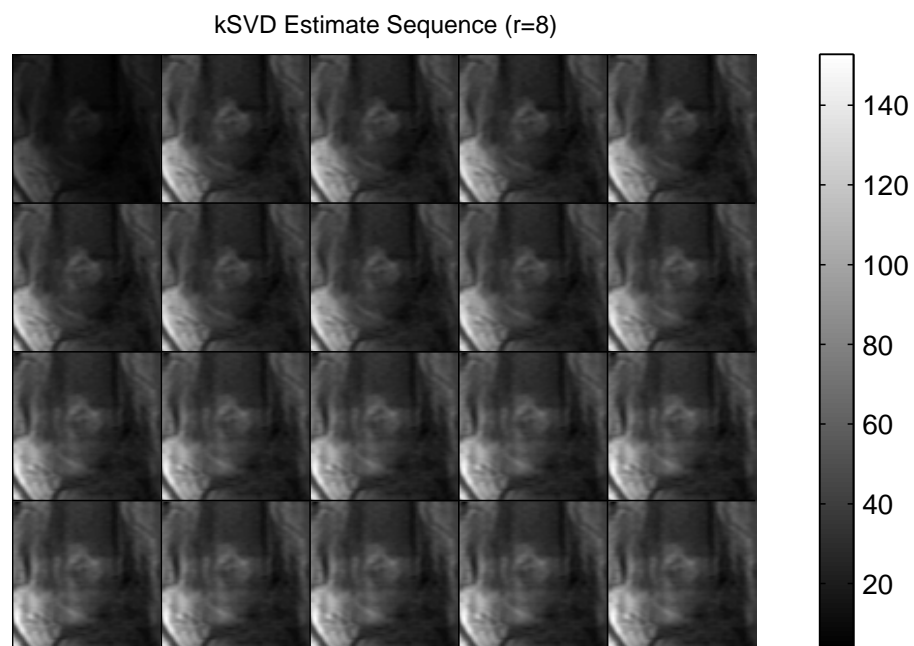


(a)

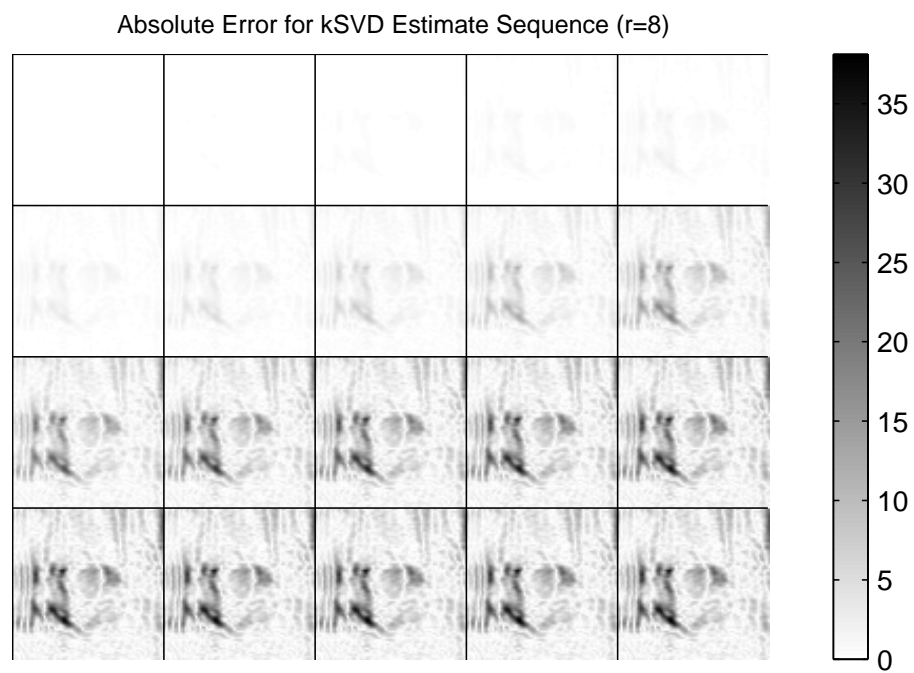


(b)

Figure 5.9: Simulated synthetic contrast change sequence acquisition using the Fourier Keyhole method: (a) Estimated Images and (b) Absolute Estimate Error



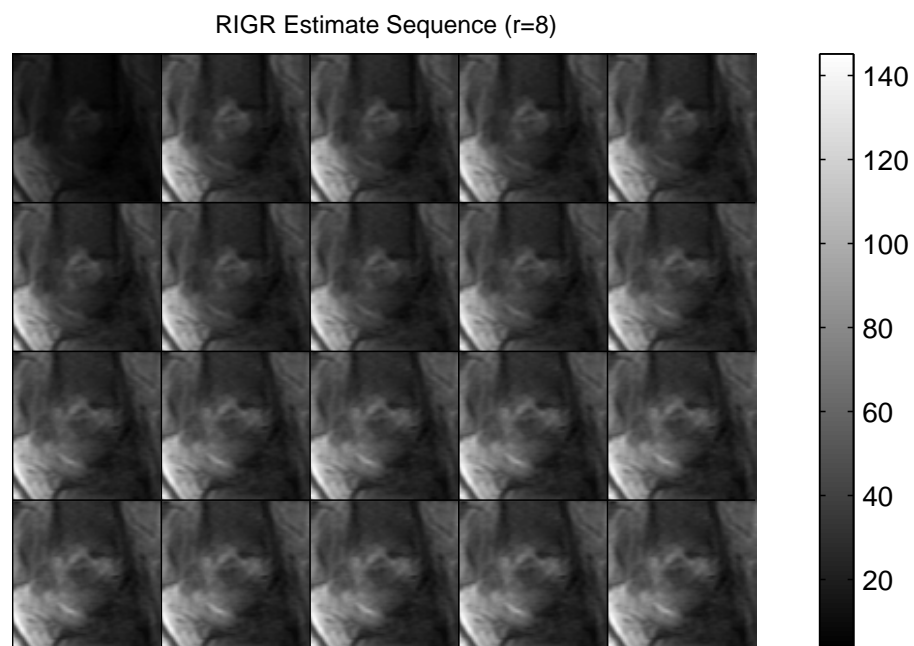
(a)



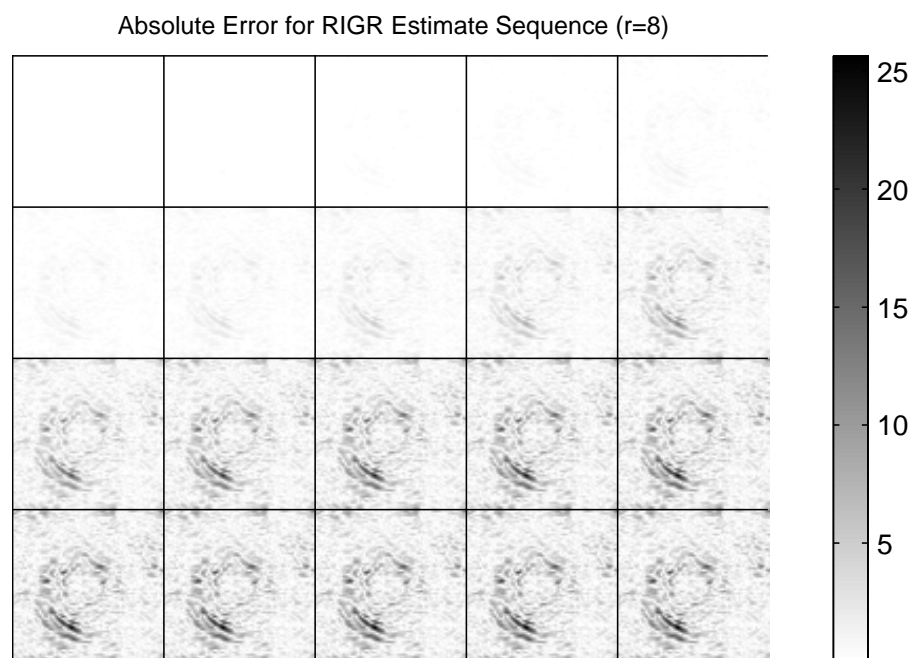
(b)

Figure 5.10: Simulated synthetic contrast change sequence acquisition using the keyhole SVD method: (a) Estimated Images and (b) Absolute Estimate Error





(a)



(b)

Figure 5.11: Simulated synthetic contrast change sequence acquisition using the RIGR method:  
 (a) Estimated Images and (b) Absolute Estimate Error

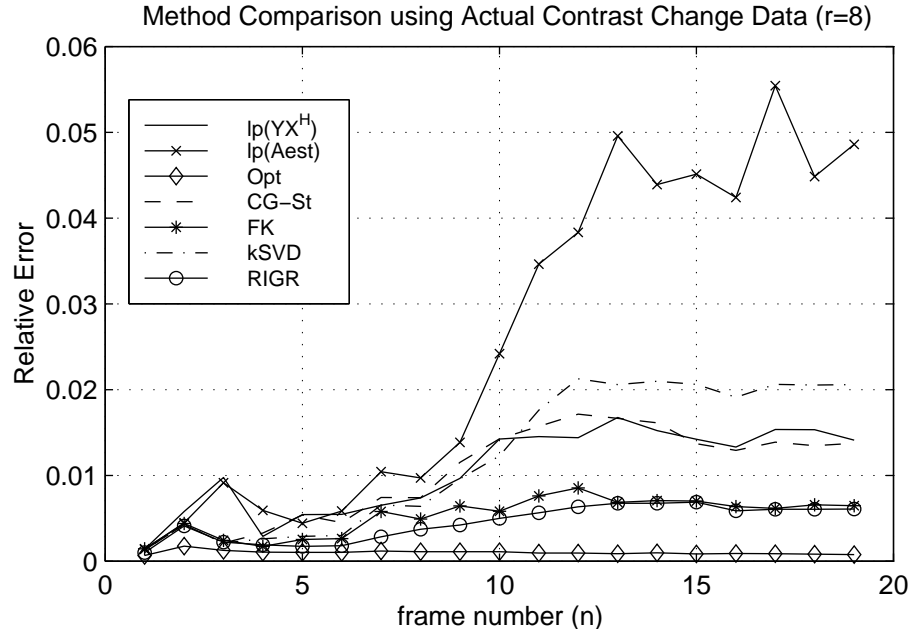


Figure 5.12: Low-order acquisition method comparison showing the relative error estimating actual contrast change MRI data

Two factors dominate the ability to track the changing images. One is the number of vectors used for each acquisition. For each of the comparisons in this section, this choice is guided by the distribution of the singular values of the reference image. A second factor is the rate of change in the image. In constructing the linear system based methods, there was an implicit assumption that the image rate of change would be much slower than the data acquisition rate.

A close inspection of Figures 5.4 shows that the relative error of the CG-St and linear predictive method  $\text{lp}(YX^H)$  decreases after frame 15 while all of the non-adaptive methods show non-decreasing error after that same frame. This implies that truly adaptive methods which choose new inputs for each image can potentially track image changes if those changes occur at a slower rate than in the first example. In a clinical setting, this would translate to more frequent low-order data acquisitions.

In the first example, the decrease in the sequence acquisition time using low-order acquisitions is nearly a factor of ten. Thus, there is certainly some flexibility to perform low-order acquisitions

at a more frequent rate. The next figure shows an experiment using the same synthetic sequence data, however the rate of acquisition is now three time faster than the original data set — or from an alternative perspective, the rate of change in the image is now three times slower. This implies that a total of  $24 = 3 \cdot 8$  vectors are applied in the same time period for the rapid acquisition sequence compared to 8 vectors in the synthetic acquisition sequence shown previously. This gives a total sequence acquisition time that is  $24/72 = 33.3\%$  as long as traditional methods which acquire the entire data set of each image — still a significant time savings.

In Matlab, the commands to generate this second set were:

```
>> d = Image34 - Image15;
>> t = atan( 0.5*([1:20]-10));
>> t_s = interp(t,3,5);
>> ImgSeq = Image15(:)*ones(size(t_s)) + d(:)* 0.5*( 1 - t./min(t_s) );
```

Figure 5.13 shows a comparison of each of the six methods for this more frequent acquisition sequence. Here again, the Fourier based methods outperform the linear system based methods. However, for  $\text{lp}(YX^H)$ , which uses the general adaptive framework with a linear prediction on  $YX^H$  to determine subsequent inputs, the image estimate does improve over time as the contrast change slows. In fact, at the end of the sequence, the estimate error is significantly lower than estimates provided by the Fourier based methods. This illustrates the fact that the  $\text{lp}(YX^H)$  adaptive method is capable of tracking changes in the image sequence if the image changes occur slowly enough.

### Synthetic grapefruit sequences

To simulate the acquisition of a dynamic MRI sequence containing bulk motion, a series of images showing a “needle” inserted into a grapefruit were acquired. The data for the 30 images in the sequence was acquired using a modified spin echo protocol. The needle motion was simulated by slowly advancing a chop-stick into the fruit, centering the fruit in the scanner, acquiring a single image, and then retrieving the fruit to advance the chop-stick slightly further to prepare for the

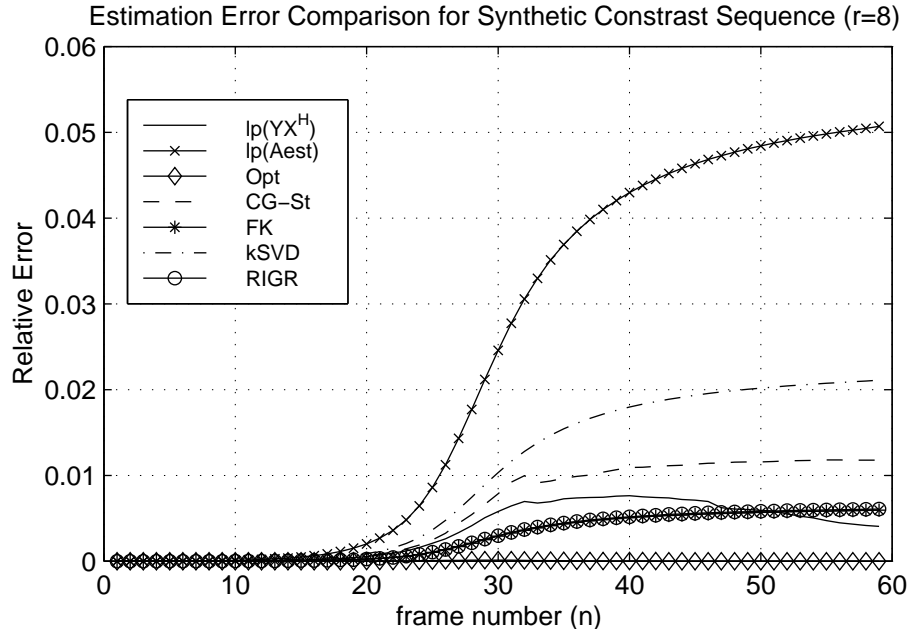


Figure 5.13: Low-order acquisition method comparison showing the relative error for the rapid acquisition synthetic contrast change sequence

next frame. This process of removing and reinserting the grapefruit reset the phase orientation of the spins in the gradient field. So while the original sequence data is complex valued, the phase component of each pixel is far more erratic from frame to frame than would be typically observed in a clinical experiment — one in which the subject stays in the scanner over the entire course of the dynamic sequence acquisition. For this reason, only the absolute value of the data was used in comparing the adaptive estimate methods. The image sequence is shown in Figure 5.19.

Close observation of the original data also revealed two secondary phenomena. First, as the chop-stick is advanced into the grapefruit it displaces some of the fluid in the grapefruit sections, pushing the juice towards the center. This is referred to as *section expansion*. The identifying feature is that some areas of low intensity in the first image become areas of high intensity later in the sequence. Second, because the grapefruit was handled in between each image acquisition, the exact location of the fruit in the magnetic field tends to shift from frame to frame. This shifting is referred to as *jitter*. Jitter is most easily identified by examining the difference between two image

frames in the sequence. Such difference images show an outline of the image structure in those image pairs where jitter occurred.

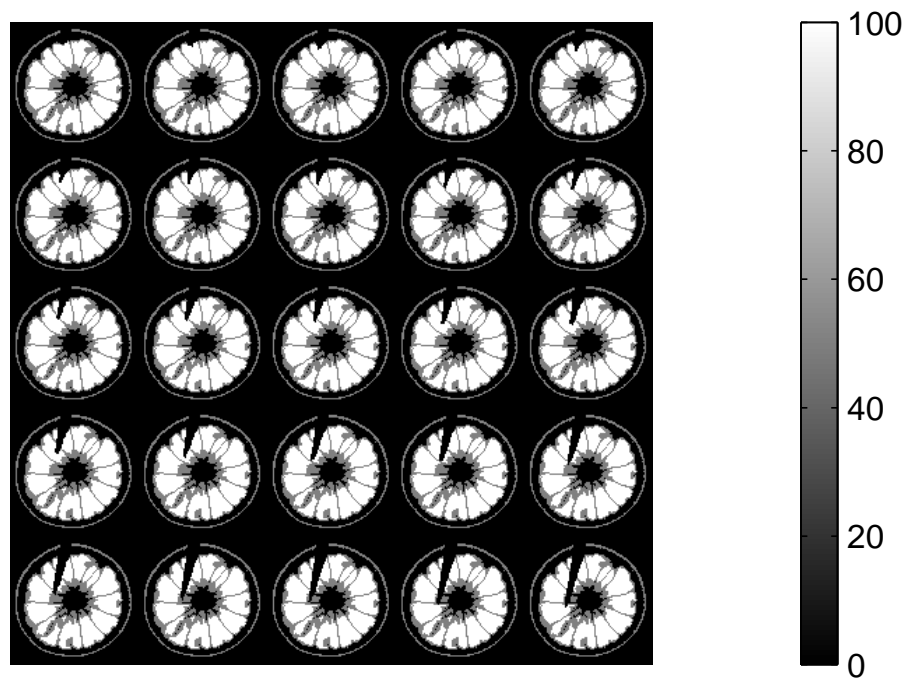
In order to isolate each of these two secondary effects, a synthetic image sequence was created. The base image of the synthetic sequence was formed by segmenting the first image of the true data sequence into three sections: a null signal region of low intensity, a juicy section of high intensity, and an interface/rind section of medium intensity. A similar segmenting procedure was applied to each of the remaining images to identify the needle motion. This needle motion data was then used to null some pixels in the base image for each frame, thus creating the synthetic sequence. Section expansion was simulated in the synthetic sequence by identifying participating pixels in the original data and then interpolating those pixels in the base image to new higher intensity values. Jitter was simulated by randomly shifting the reference points of the synthetic data relative to the field of view.

With each of these three phenomena isolated (needle motion, section expansion, and jitter), multiple synthetic sequences could be created. To more closely approximate the acquisition of the real data, noise could also be added. The examples below show the relative error comparison for the synthetic cases given in Table 5.3 for both the noise-free and additive noise cases.

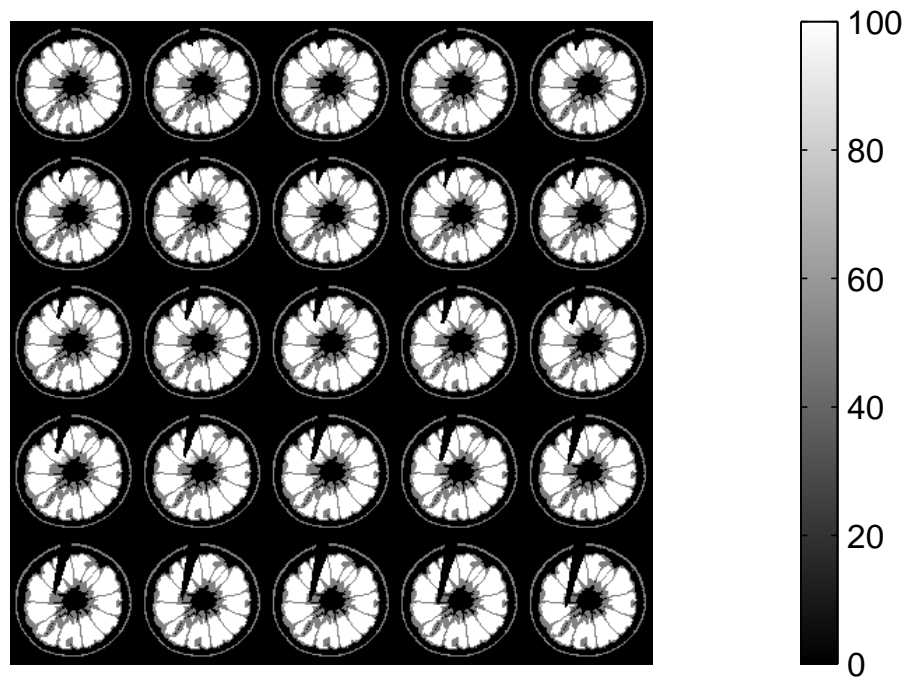
Sequence	Original Images	Relative error comparison
Basic chopstick motion:	Figure 5.14(a)	Figure 5.15
Section expansion sequence:	Figure 5.14(b)	Figure 5.16
Sequence with jitter:	-	Figure 5.17
Section expansion and jitter sequence:	-	Figure 5.17

Table 5.5: Table of synthetic test sequences and associated figures

The basic synthetic sequence is shown in Figure 5.14(a). A comparison of the relative error for each of the seven methods tested is given in Figure 5.15. In both the case additive noise and noise free cases, RIGR provided the best realizable estimates. In fact, this synthetic experiment closely mirrors the original examples given in [13], so this result is not surprising. Of greater interest in this figure is that the relative error of the adaptive methods (CG-St,  $\text{lp}(YX^H)$ , and  $\text{lp}(\text{Aest})$ )



(a) Basic Sequence



(b) Section Expansion

Figure 5.14: Original synthetic grapefruit sequences

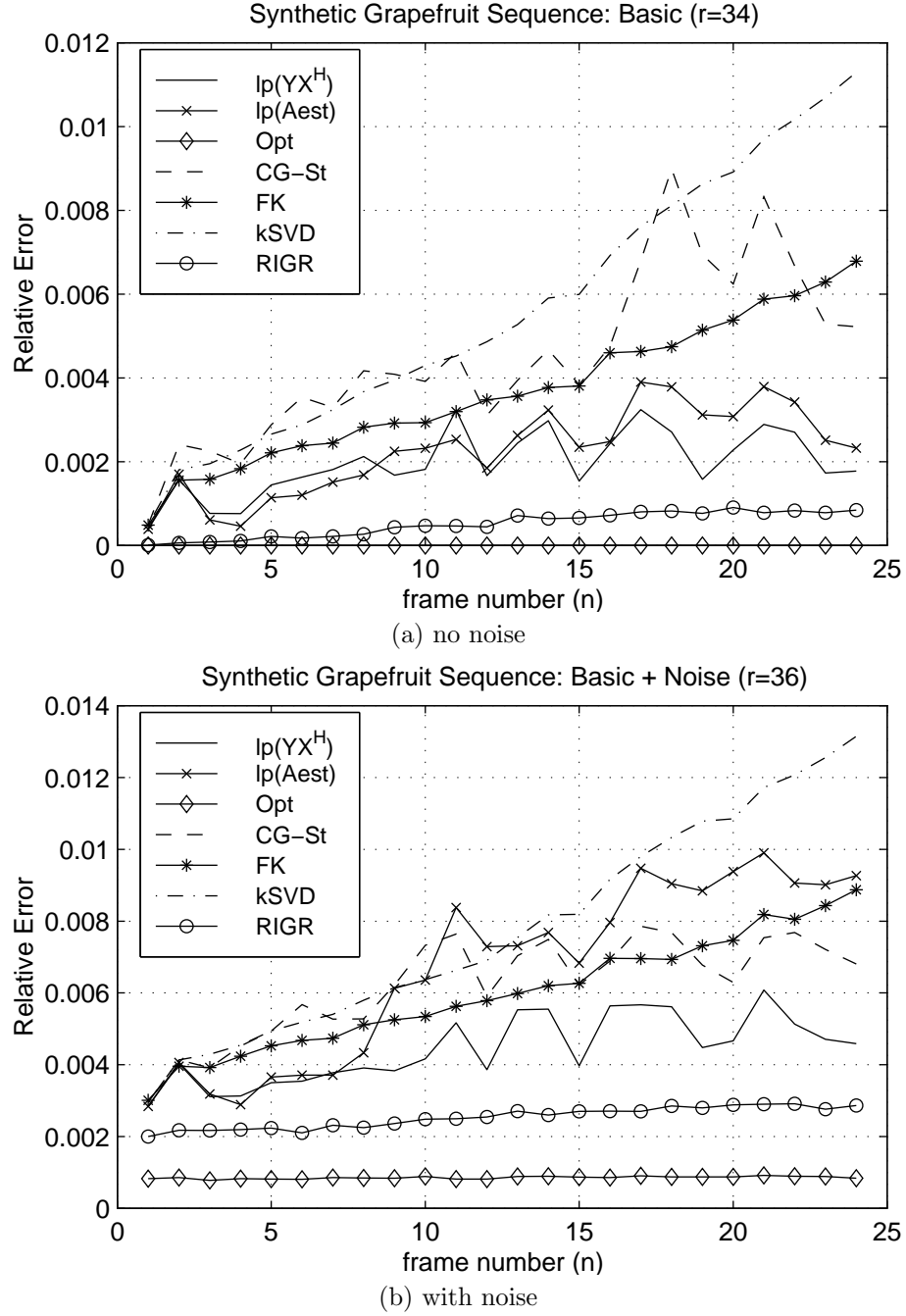


Figure 5.15: Relative error comparison for basic synthetic grapefruit sequence

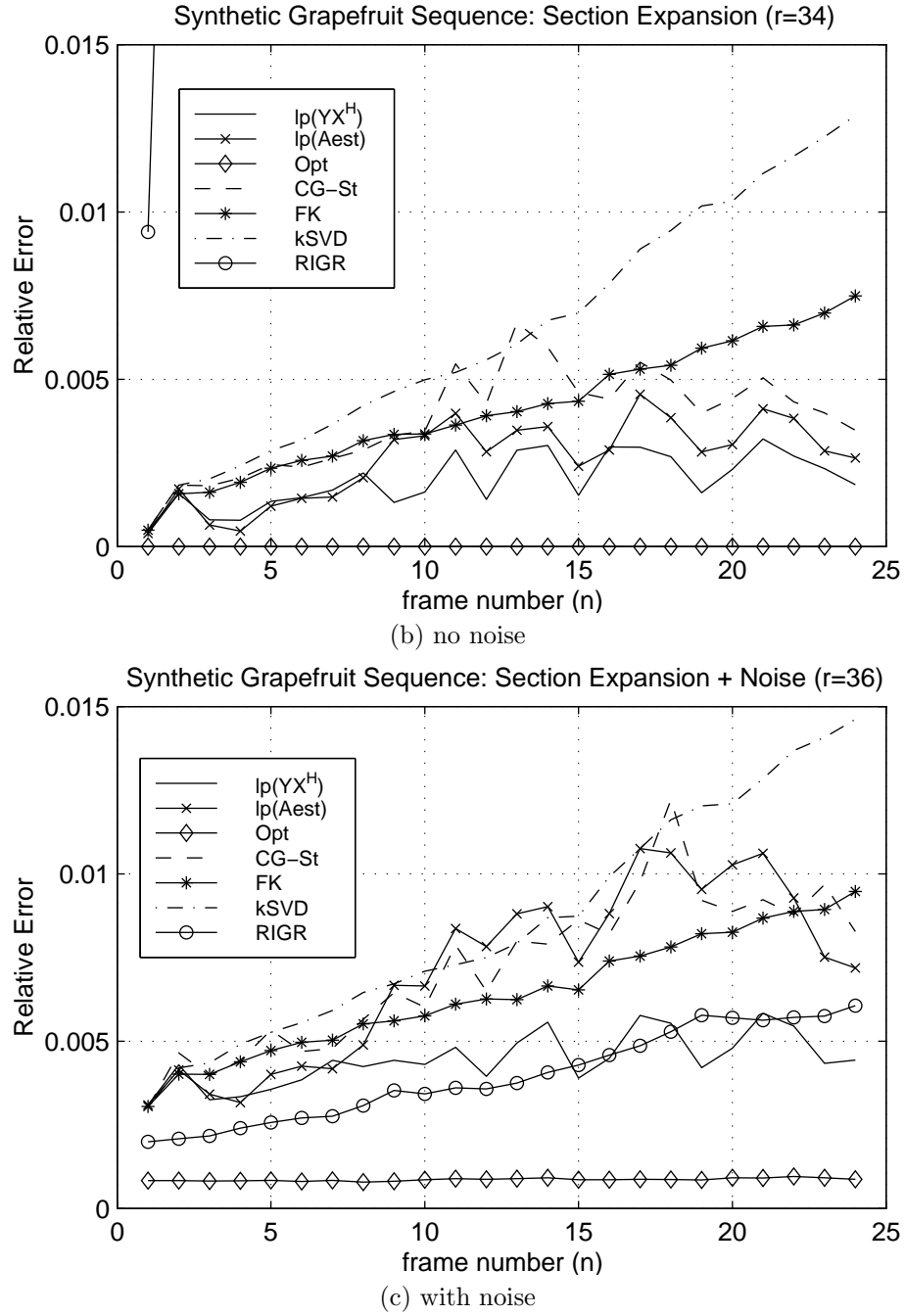


Figure 5.16: Relative error comparison for synthetic grapefruit sequence with section expansion



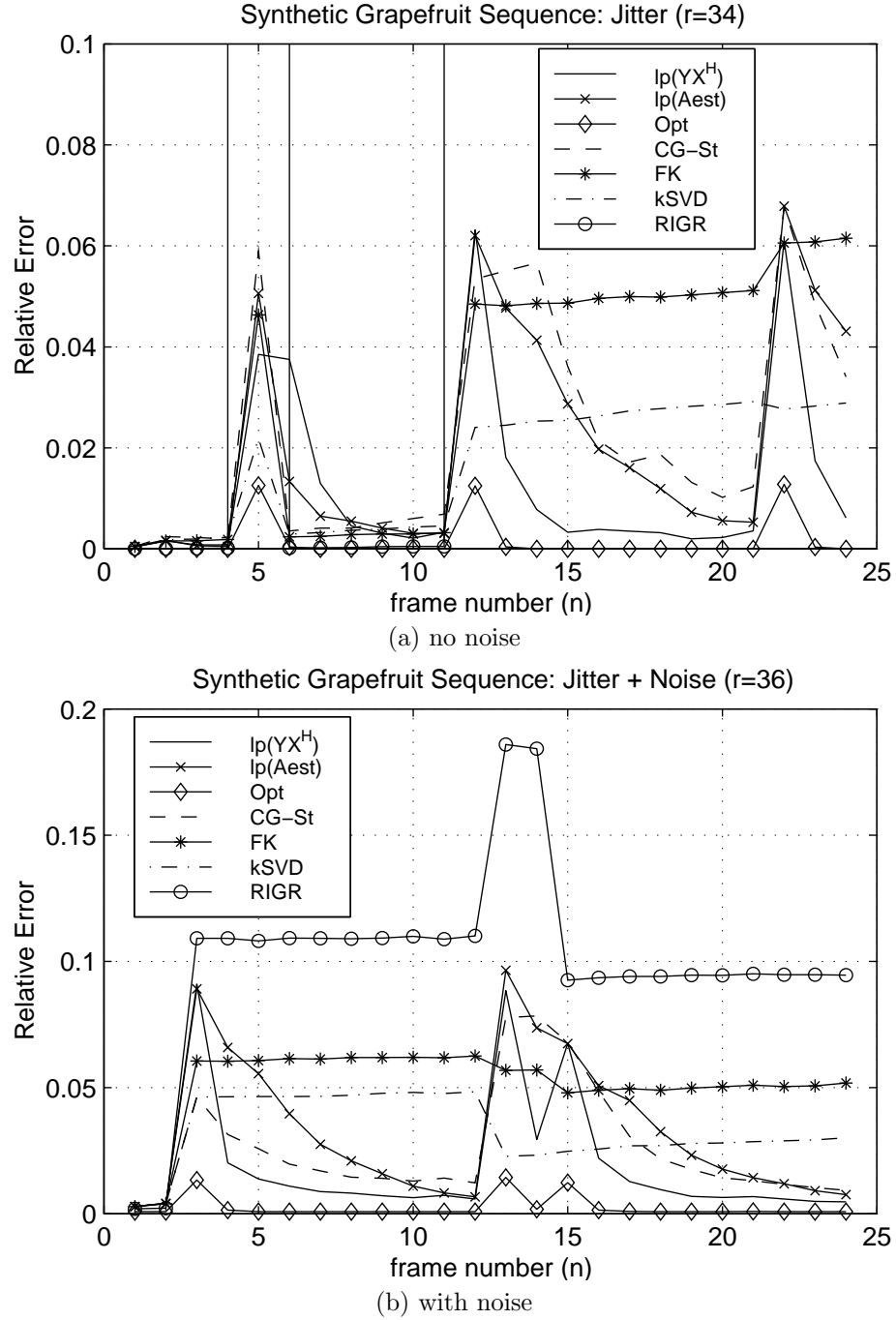


Figure 5.17: Relative error comparison for synthetic grapefruit sequence with random jitter

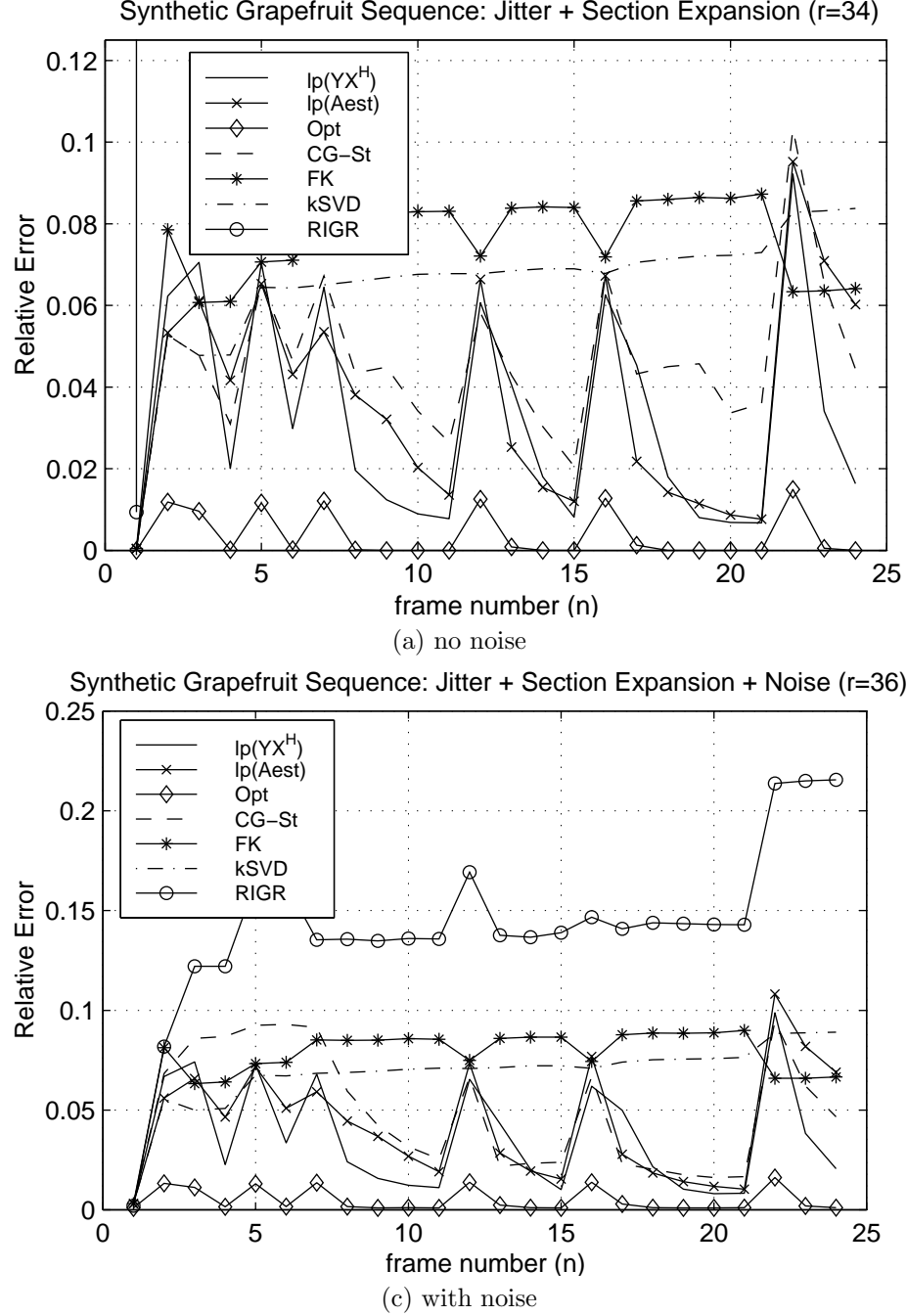


Figure 5.18: Relative error comparison for synthetic grapefruit sequence with section expansion and random jitter

shows wide variability over the course of the sequence. This is in contrast to the steadily increasing error of the SVD, FK, and RIGR methods. This shows that the adaptive methods are able to correct errors in the image estimate for motion sequences.

The synthetic section expansion sequence is shown in Figure 5.14(b). The significant feature of this sequence is that fluid is pushed towards the center of the grapefruit as the chopstick cuts deeper into the fruit. Thus a number of pixels show a change from low intensity to high over the course of the sequence. A comparison of the relative error for each of the seven methods tested is given in Figure 5.16. In this figure we note the RIGR is unable to track this type of change at all in the noise free case. This is because the spatial envelope component of the RIGR image estimate restricts the ability to model increasing intensity contrast changes. In the additive noise case, the spatial envelope has more flexibility and RIGR is able to track the image changes. However, the new  $\text{lp}(YX^H)$  method provides lower reconstruction error towards the end of the sequence. The distribution of the other image estimation methods are consistent with the basic sequence in Figure 5.15.

The synthetic jitter sequence is not shown here because a shift in the image by one pixel is difficult to observe on a printed page. A comparison of the relative error for each of the seven methods tested on the jitter sequence is given in Figure 5.16. In this figure the discontinuity points in each relative error curve clearly identify the occurrence of jitter in the sequence. For this sequence type, the spatial envelope that RIGR imposes prevents that method from tracking the image shifts. From the relative error curve, we see that RIGR has an extremely difficult time tracking jitter. The other non-adaptive methods show a relatively constant level of error between each jitter occurrence. In sharp contrast, the new methods all correct for jitter, showing a steep convergence curve that approaches the theoretically optimal estimate within ten jitter-free frames.

Figure 5.18 shows a relative error comparison for a sequence containing all of the independent features simulated in this section: the basic chopstick motion, section expansion, and jitter. One would expect that with the addition of noise this synthetic sequence would be most similar to

clinical sequence acquisition experiments. We see again that section expansion and jitter cause estimation failure in RIGR, which goes off the chart in the noise-free case. The SVD and FK methods show the next level of estimate quality, providing relatively constant error performance over the course of the sequence. However, the adaptive methods (CG-St,  $\text{lp}(YX^H)$ , and  $\text{lp}(\text{Aest})$ ) show again decreasing error after each occurrence of jitter in the sequence. The jitter occurs in this sequence with enough frequency that the adaptive method's estimate quality never reach the quality provided by the optimal method. The improvement over the static input methods is substantial however.

### Acquisition simulations using actual MRI data

The following figures compare each of the methods discussed above for a simulated dynamic MRI sequence acquisition using actual MRI data. The dynamic sequence was created by acquiring images of a chopstick as it was inserted into a grapefruit. A more complete description of the data acquisition method is given in the first paragraph of Section 5.3 and the original data sequence is shown in Figure 5.19. Figures 5.21 through 5.27 show image estimates constructed using each of the six methods given in Table 5.3.

Acquisition method	Acronym	Estimates
Theoretically Optimal	[ Opt ]	Figure 5.21
Linear prediction with input $\check{A} = YX^H$	[ $\text{lp}(YX^H)$ ]	Figure 5.22
Linear prediction with input $\check{A} = YX^H + \hat{A}_{n-1}(I - XX^H)$	[ $\text{lp}(\text{Aest})$ ]	Figure 5.23
Conjugate gradient on Stiefel manifold	[ CG-St ]	Figure 5.24
Fourier keyhole	[ FK ]	Figure 5.25
Keyhole SVD	[ kSVD ]	Figure 5.26
RIGR	[ RIGR ]	Figure 5.27

Table 5.6: Methods and associated figures used in simulated acquisition using actual MRI data

Each figure shows both the image estimate and the absolute estimate error. Figure 5.20 compares the relative error,  $\text{re}(\hat{A}_n) = \|A_n - \hat{A}_n\|_F^2 / \|A_n\|_F^2$ , for the seven methods.

Clearly, the adaptive methods developed in this work perform much better than the methods

previously proposed. Each of the non-adaptive methods (FK, kSVD, RIGR) tend to have similar relative performance although the apparent errors are unique to each method. From Figure 5.26, the SVD method shows significant “blocky” artifacts that appear at the edge of the fruit. Figure 5.27 shows that RIGR is unable to properly estimate those regions that contained a low signal level in the first image. Specifically, a significant amount of error occurs in those grapefruit sections where the fluid has been pushed towards the center of the fruit by the chopstick. As expected, the Fourier Keyhole method suffers from significant blurring. This is especially apparent along the edges of the chopstick. All of these methods grow progressively worse over the course of the sequence. This is because either the input vectors (in the case of the SVD method) or the image estimate model (in the case of FK and RIGR) are closely biased to the original image. Thus these three methods are unable to correct for slight shifts in the location of the fruit.

From the relative error comparison of Figure 5.20, the linear predictive and CG-St methods show significantly lower error than the other three methods. The linear prediction method,  $\text{lp}(YX^H)$ , performs better than the CG-St method, showing results consistent with the synthetic sequences examined above. However, the vast difference between the physically realizable methods and the theoretically optimal at this order of reconstruction show that there is significant room for improvement in the area of low-order dynamic MRI.

## 5.4 Summary of the dynamic problem

This section presented new methods for estimating dynamic MRI sequences. A general adaptive image estimation framework was developed. Analysis showed that previous low-order estimation methods, specifically Fourier Keyhole and SVD methods, are special cases of this general framework. This section also explored issues related to the dynamic determination of input vectors. This analysis identified that methods to choose new vectors based on previous image estimates tend to be biased towards the input vectors used to reconstruct those estimates. We denote this tendency

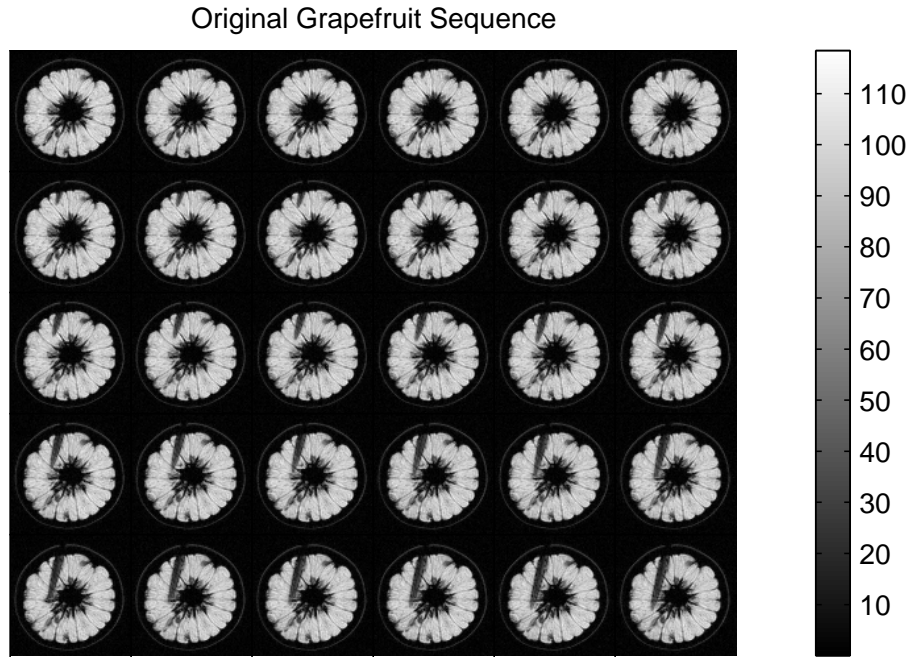


Figure 5.19: Original image sequence for simulated grapefruit acquisition

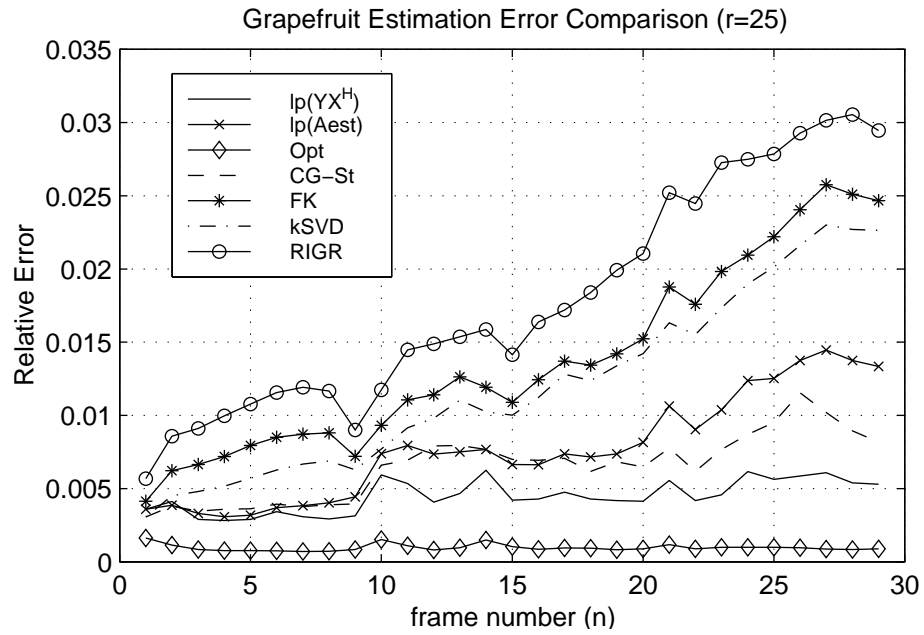
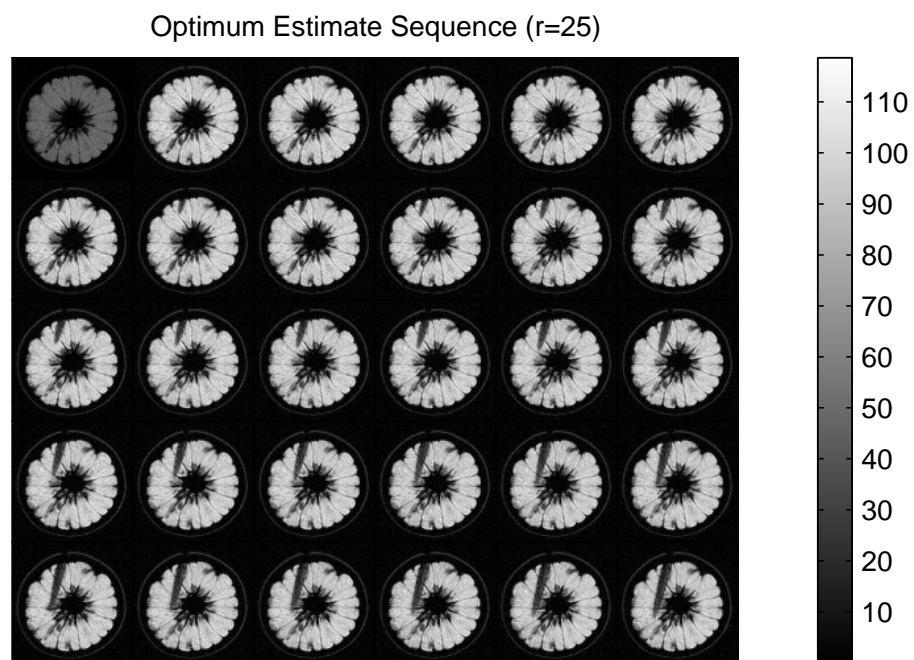
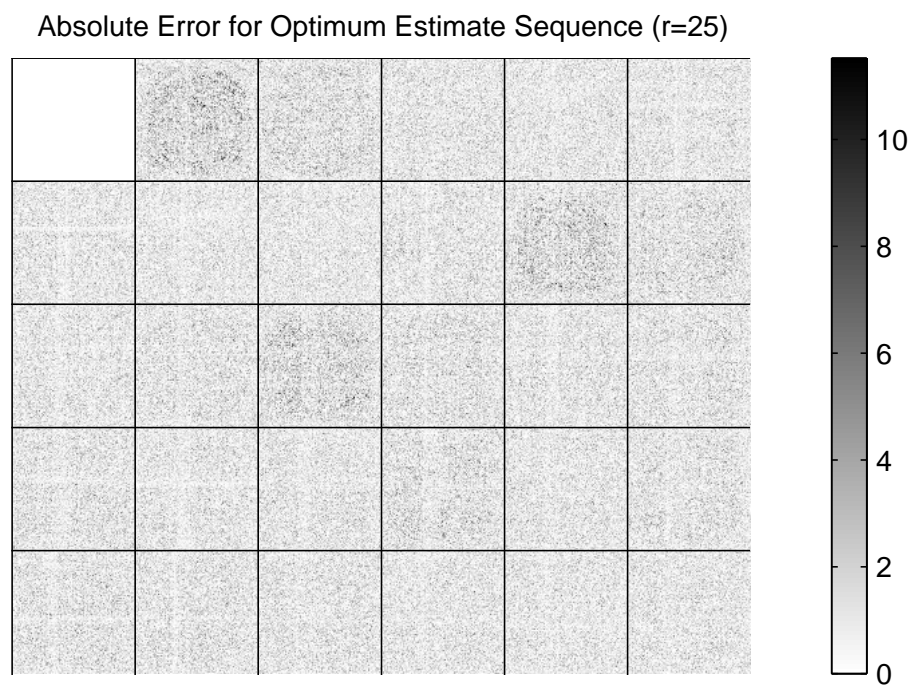


Figure 5.20: Relative error comparison for simulated grapefruit sequence acquisition

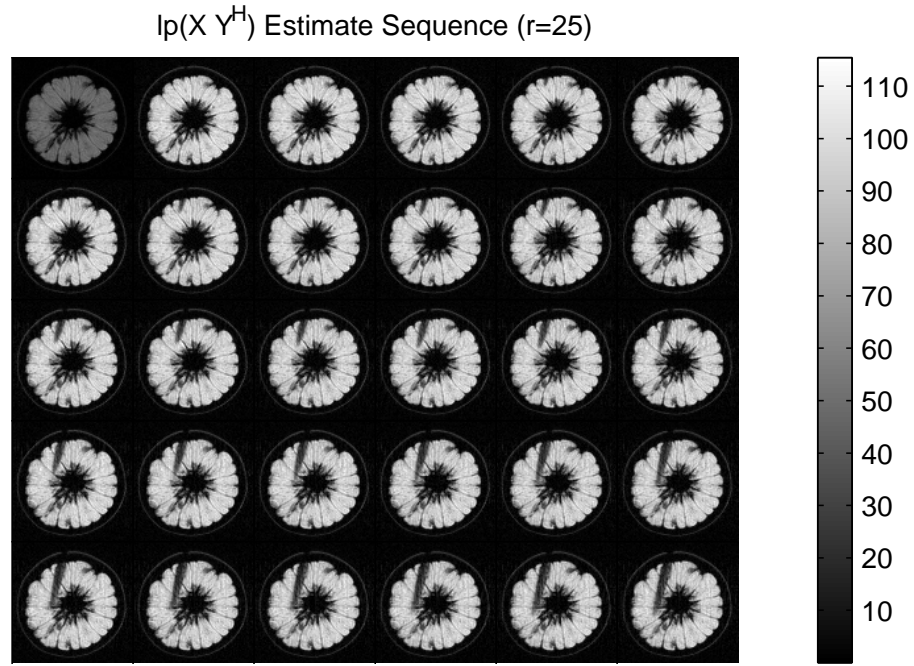


(a)

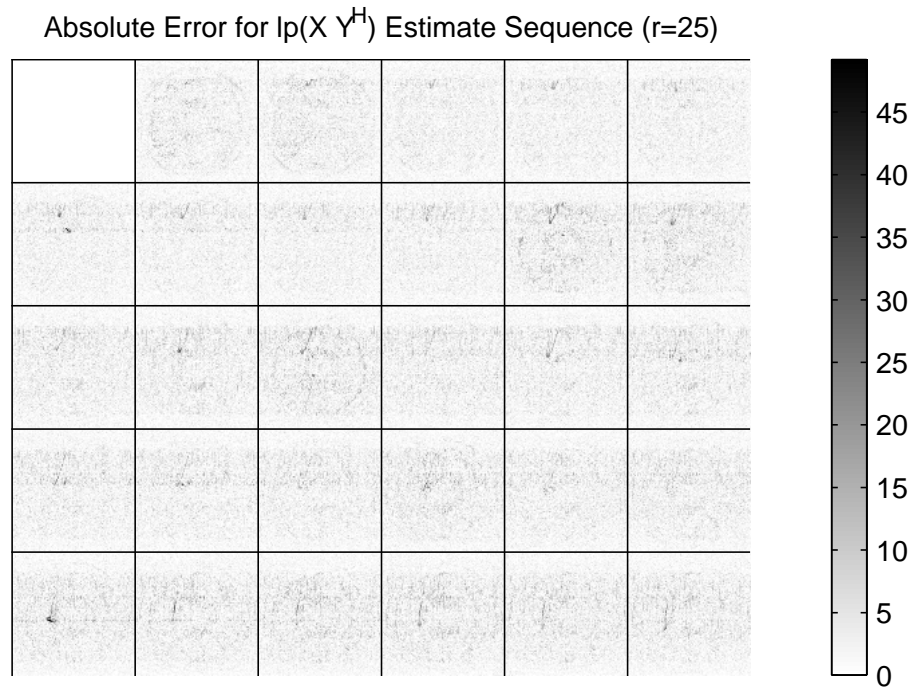


(b)

Figure 5.21: Simulated grapefruit sequence acquisition using Optimal method: (a) Estimated Images and (b) Absolute Estimate Error



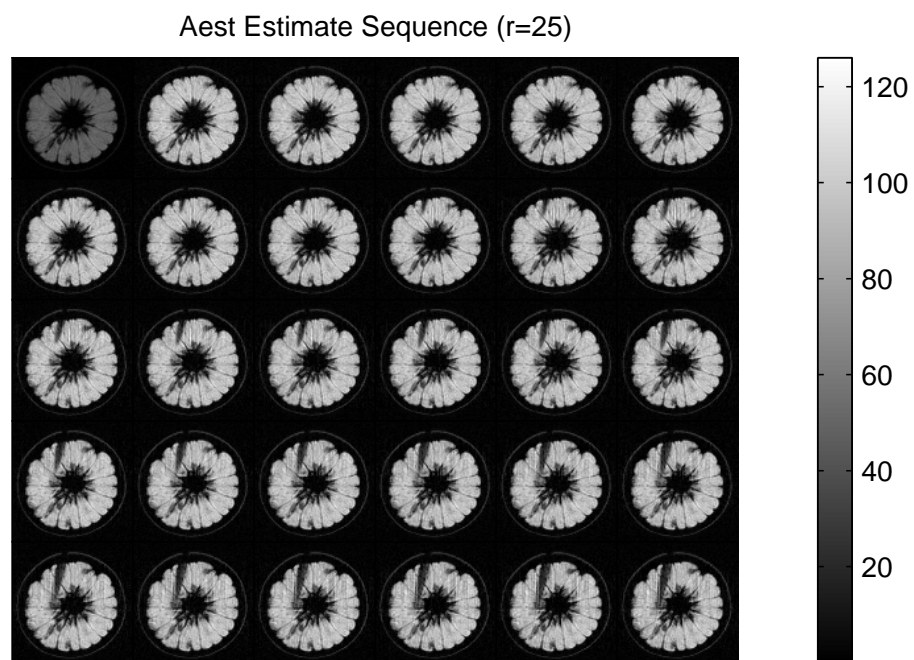
(a)



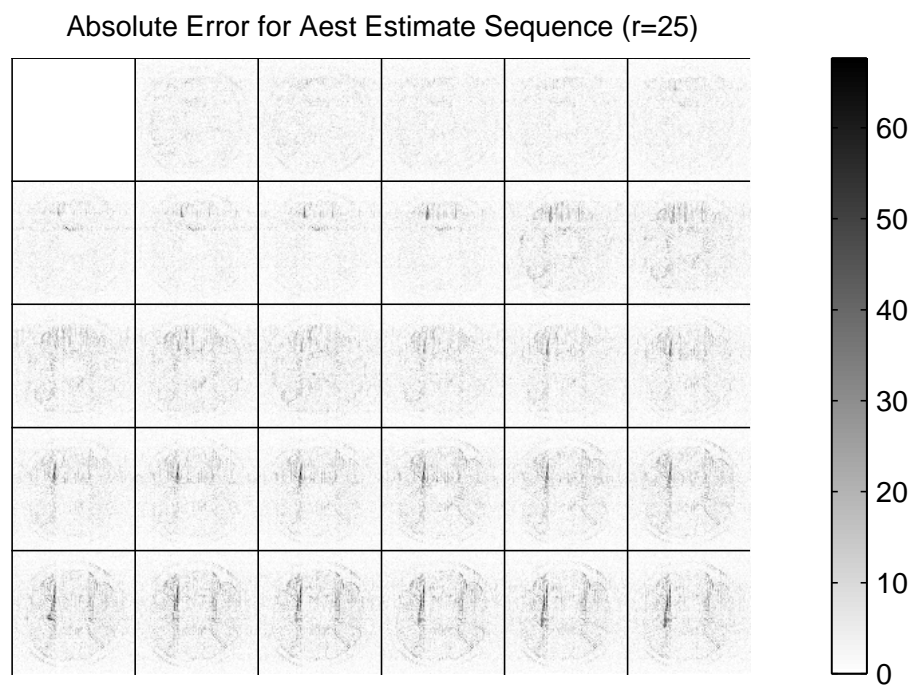
(b)

Figure 5.22: Simulated grapefruit sequence acquisition using Linear Predictor  $[\text{lp}(Y X^H)]$  Method:  
 (a) Estimated Images and (b) Absolute Estimate Error



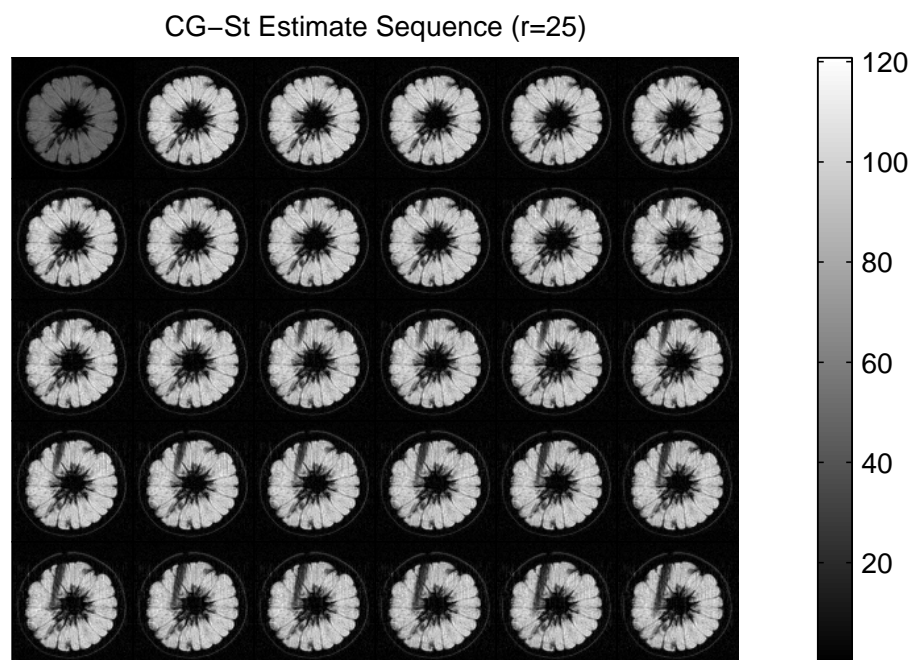


(a)

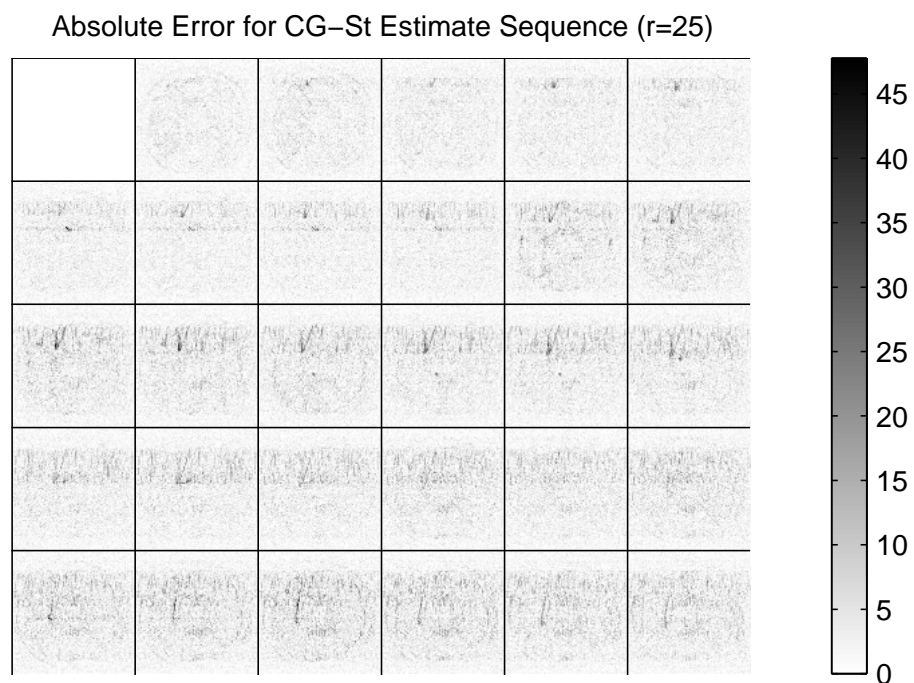


(b)

Figure 5.23: Simulated grapefruit sequence acquisition using Linear Predictor  $[lp(Aest)]$  Method: (a) Estimated Images and (b) Absolute Estimate Error

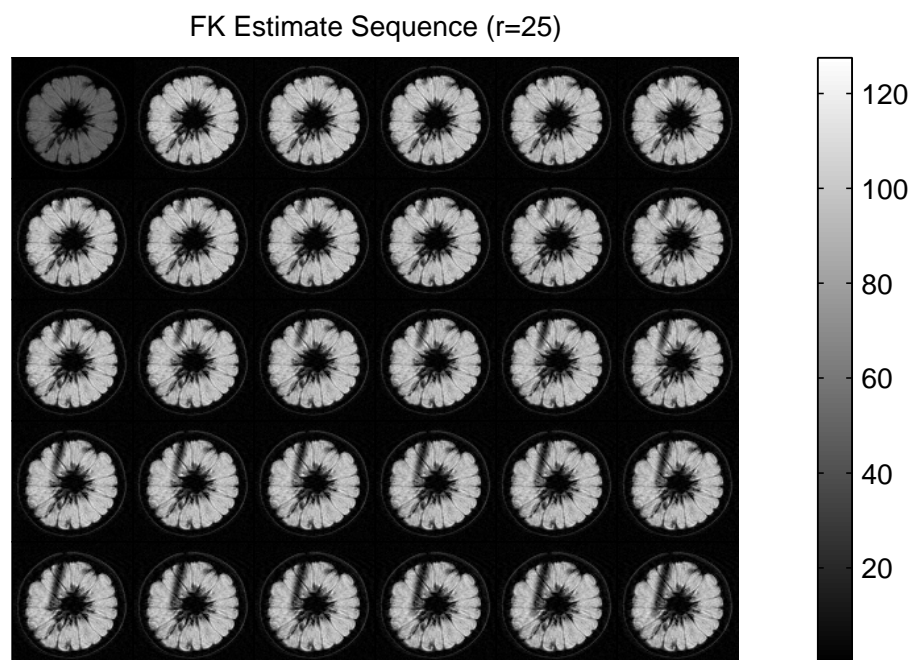


(a)

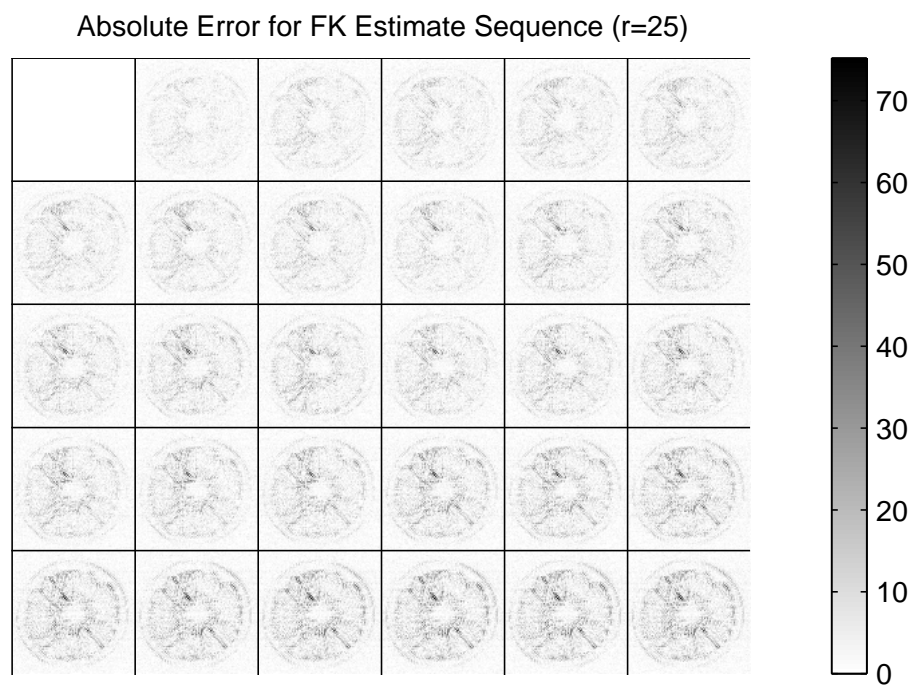


(b)

Figure 5.24: Simulated grapefruit sequence acquisition using CG-St method: (a) Estimated Images and (b) Absolute Estimate Error

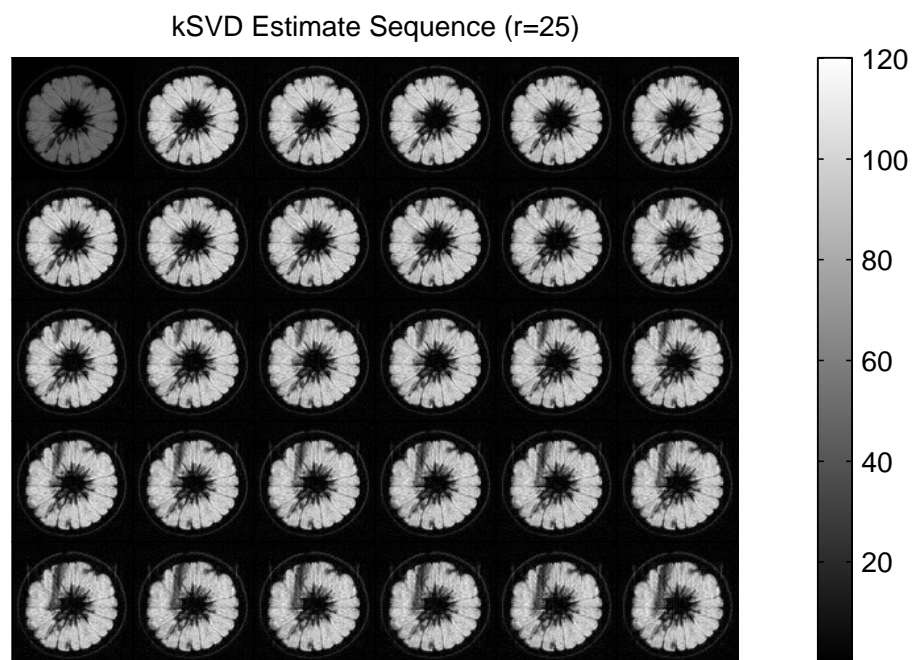


(a)

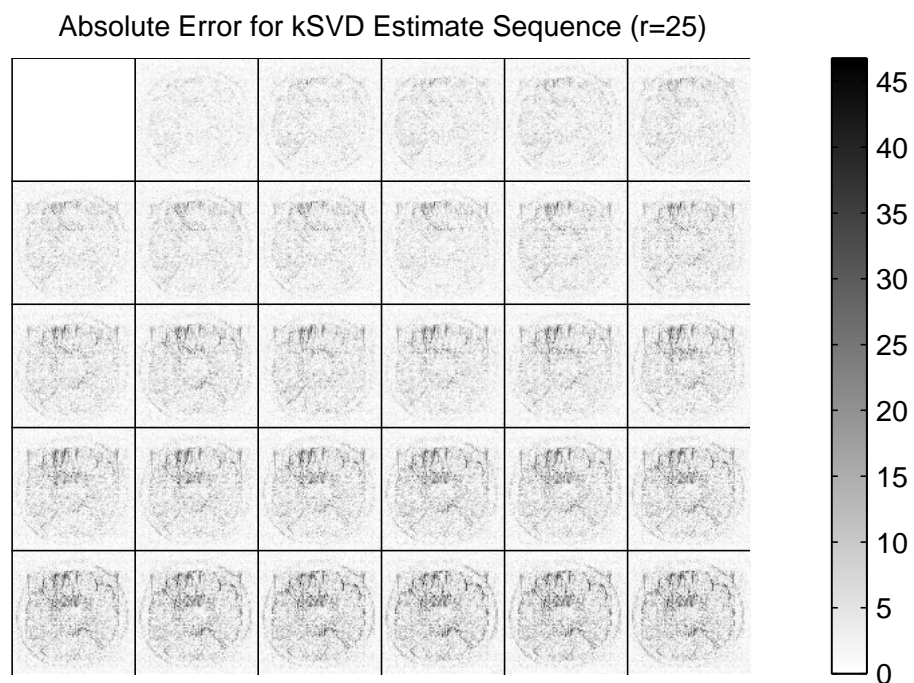


(b)

Figure 5.25: Simulated grapefruit sequence acquisition using Fourier Keyhole method: (a) Estimated Images and (b) Absolute Estimate Error

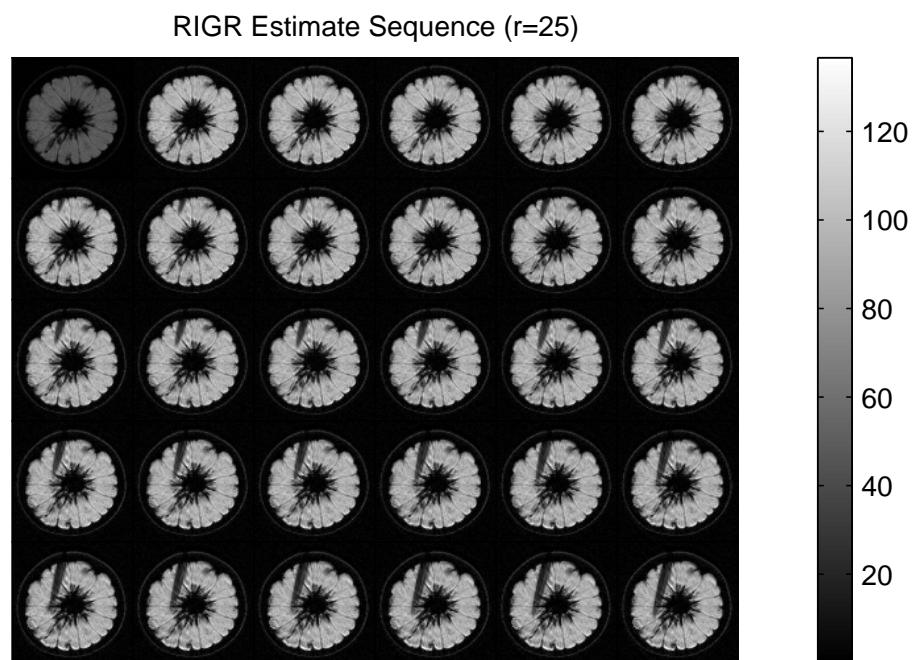


(a)

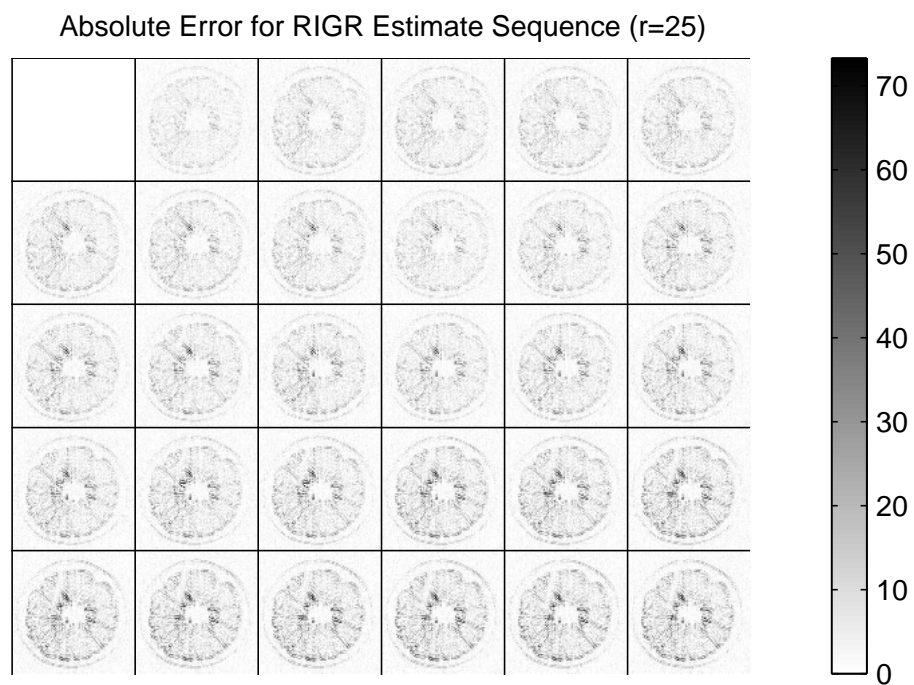


(b)

Figure 5.26: Simulated grapefruit sequence acquisition using keyhole SVD method: (a) Estimated Images and (b) Absolute Estimate Error



(a)



(b)

Figure 5.27: Simulated grapefruit sequence acquisition using RIGR method: (a) Estimated Images and (b) Absolute Estimate Error

as the subspace trap. Two input identification algorithms were presented which overcome the subspace trap. These methods used the conjugate gradient method on the Stiefel manifold, CG-St, and linear temporal prediction, lp( $\cdot$ ), respectively. The linear prediction method has the added feature of input flexibility. Specifically, predictor inputs may be different from the image sequence estimates. Two such inputs were identified and compared in this section: the full rank estimate,  $\hat{A}_n$ , and the low-rank instantaneous estimate,  $YX^H$ . Examples in this section show that given a sufficient number of input vectors and rapid enough data acquisition rate, these adaptive methods provide image estimates that exhibit a significant quality improvement over previous low-order dynamic sequence estimation methods.

## Chapter 6

# Conclusions and future research

This work showed that signal processing techniques can be applied to the acquisition of MRI images to improve the efficiency of dynamic sequence acquisitions. The methods presented draw primarily from the linear system model of Panych and Zientara, and thus are applicable with imaging modalities for which the model holds true.

The significant results of the static problem work was the CCD algorithm, a numerical method to efficiently represent an arbitrarily shaped region of interest in a static image. While similar to traditional matrix completion problems, this new method provides a numerical solution that does not impose any inherent structure on the matrix to be completed. However, as shown in Section 4.4.2, the utility of this method for acquiring MRI images is somewhat limited. The presence of noise in the image acquisition process severely corrupts the ability of this method to provide high quality estimates of the ROI. A discussion of how to possibly overcome noise in the image acquisition and reconstruction process is discussed below in Section 6.2.

The significant results of the dynamic problem were the development of a general adaptive estimation framework and two methods to dynamically determine suitable acquisition inputs. Together, these new methods address both the image estimation and input identification aspects of the dynamic problem. Analysis of the adaptive framework shows that the Fourier Keyhole and

SVD methods of low-order acquisition are special cases of the general method presented in Section 5.1. With the inclusion of the input identification methods, both the conjugate gradient on the Stiefel manifold (CG-St) method and the linear predictor methods ( $\text{lp}(YX^H)$  and  $\text{lp}(\text{Aest})$ ), this truly adaptive estimate method outperforms previously proposed low-order acquisition methods for a variety of dynamic MRI sequence acquisition simulations. The primary open question is: Do other realizable methods exist that more closely approach the performance of the theoretically optimal method? Possible answers to this question are discussed in Section 6.2.

It should be noted that the general adaptive framework presented in Chapter 5 is applicable to any dynamic system estimation problem where one has control over the system excitation used during data acquisition. This includes a variety of imaging modalities including sonar, radar, and possibly X-ray / CT imaging. The work presented in this thesis also provides a framework to determine the most appropriate set of excitation vectors to use for such imaging modalities.

## 6.1 Open static problem questions

The primary conclusion from the static work is that while inputs can be designed to efficiently reconstruct an arbitrarily shaped region of interest, reconstruction quality using these vectors in a noisy environment is limited. Thus, any future work focused on the static problem must resolve the presence of noise in the system.

One may start by recasting the linear system model as

$$Y = AX + \mathcal{N} \tag{6.1}$$

where  $\mathcal{N}$  is a noise matrix identical in size to  $X$  and  $Y$ . My examination of the experimental output data indicates that the noise is uncorrelated and normally distributed. Thus  $\mathcal{N}$  can be described by a Gaussian probability distribution.

Introducing the system model with noise into the static ROI reconstruction problem of Chap-



ter 4, the cost function becomes

$$\mathcal{J} = \|S \circ (A - (AX + \mathcal{N})L^H)\|_F^2 = \|S \circ (A - AXL^H + \mathcal{N}L^H)\|_F^2. \quad (6.2)$$

Close examination of (6.2) identifies two conflicting design constraints. First, the norm of  $X$  (which translates to power in the applied rf excitation signal) must be kept close to unity in order to keep the linear system model valid. Second, the norm of  $L$  must be small in order to keep the noise matrix from dominating the reconstruction.

However, a third constraint on the problem is introduced by the scanner hardware. Specifically, the maximum amplitude of the input rf vectors is '1'. If the singular value spectrum of  $XL^H$  is applied solely to  $X$ , then the combined effect of both hardware scaling and spectrum scaling is to reduce the amplitude of a significant number of the input vectors. For the experiments shown in Section 4.4.2, the columns of  $X$  associated with the large singular values of  $XL^H$  did not generate a strong signal. In fact, the columns of  $X$  associated with the large singular values of  $XL^H$  may be in the range space of the image slice. Thus, there is no guarantee that these vectors will produce any measurable signal at all. In addition, if the singular value spectrum of  $XL^H$  is applied to  $X$ , the vectors of  $X$  associated with the smaller singular values may be significantly reduced and not produce a measurable output signal either.

Thus, it appears that  $X$  and  $L$  must both be orthogonal for most practical MR image acquisitions. A simple strategy is to constrain  $L = X$  and minimize (6.2) while constraining  $X$  to the Stiefel manifold, i.e.,  $X \in St(n, r)$ . However, preliminary results show that this solution path is not beneficial. For a random matrix  $A$  of size  $15 \times 15$ , and the diamond shaped selection matrix  $S$  shown in Figure 6.1, we find that such a constrained problem produces a solution with significantly higher error than the original CCD algorithm provides. Table 6.1 shows the statistics of applying both the CCD algorithm and a minimization of (6.2) with  $X = L$  and  $X \in St(n, r)$  to 100 random matrices  $A$ . The elements of  $A$  were Gaussian distributed with a mean value of zero and a standard deviation of one. These results are also compared with the value of (6.2) using the

right singular vectors of  $A$  and the right singular vectors of  $S \circ A$ , denoted  $\text{rSV}(A)$  and  $\text{rSV}(S \circ A)$  respectively. Clearly, there is no significant advantage provided by minimizing (6.2) with  $X = L$  and  $X \in \text{St}(n, r)$ . The performance is nowhere near the performance of the CCD algorithm, and is only slightly better than using the right singular vectors of  $S \circ A$ .

	CCD	$X \in \text{St}(n, r)$	$\text{rSV}(A)$	$\text{rSV}(S \circ A)$
mean	0.2996	0.4646	0.5848	0.4742
standard deviation	0.0438	0.0436	0.0445	0.0432

Table 6.1: Error comparison between CCD and  $X = L \in \text{St}(n, r)$  methods for 100 random matrices

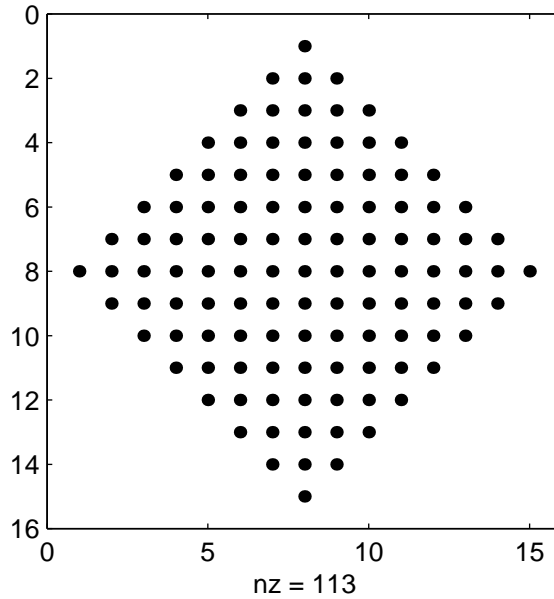


Figure 6.1: Diamond shaped region of interest

Thus, from this perspective, the best strategy may be to redesign the CCD algorithm with the constraint that both  $X$  and  $L$  lie in the Stiefel manifold. However, while possible, I highly doubt that this method will match the results of the unconstrained CCD algorithm.

## 6.2 Open dynamic problem questions

The research on the dynamic problem provided a general estimation framework and showed two methods for dynamically determining input vectors that provide superior MR image estimates over the current state of the art. In sharp contrast to previous methods which focused on the acquisition of contrast change sequences, these new methods are particularly well suited to handle many types of image changes including bulk motion and jitter. Thus, it is expected that these methods are quite well suited to handle clinical imaging scenarios.

However, comparing the image quality of the realizable low-order methods with the theoretically optimal method, we see that there is significant room for improvement in the realizable methods. For example, the image estimate error in theoretically optimal estimates appears to be uncorrelated with the structure of the original image. This is not the case for the realizable methods. Specifically, errors often appear at the edges of regions showing motion change. Thus the open question discussed in this section is, How can one determine realizable inputs that are close to the theoretically optimal?

The crux of this problem is somewhat obscure. The basic concept of the prior SVD methods is to track the images and determine new inputs from the previous estimates. We showed in Section 5.2.1 that this strategy is somewhat limited due to the bias of future inputs to previous inputs. A second point of view is to track the dominant subspace of the underlying image system. In general, this is very difficult to do. Consider an example given by Stewart and Sun in [40]. The right singular vectors of the matrix

$$\begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix} = UDV^H \quad \text{are} \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

In the limit that  $\epsilon \rightarrow 0$ , the right singular vectors of the matrix become

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This example shows that the transition of the right singular vectors of a matrix can be discontinuous even for very small changes in the matrix itself. However, keep in mind that in the adaptive framework the theoretically optimal method uses *differences* between matrices to identify suitable input vectors. Generally speaking, there is little reason to suspect that the dominant singular vectors of difference matrices are at all similar over the course of a dynamic MRI sequence. Thus, we conclude that tracking singular vectors is of limited utility.

Of the methods examined thus far, the linear predictor methods have the greatest potential for approaching the estimate quality provided by the theoretically optimal method. Two main avenues of research are available. First, the linear predictor methods presented in this thesis used temporal prediction on a pixel-by-pixel basis. Given the fact that bulk motion changes so strongly affect the estimate quality, it is reasonable to expect that including spatial changes in the predictor will provide superior performance. Second, incorporating image change models will likely improve the image estimate quality. These models would likely need to be case specific such as a temporal model matched to the stimulus used in functional MRI studies or a periodic bulk motion model to predict changes introduced by breathing or cardiac activity.

There are other questions of secondary priority that remain open at this juncture. These relate to specific design criteria in the acquisition of clinical dynamic sequences. For example, currently there is no clear guideline for selecting the number of input vectors to use for each image acquisition/reconstruction. The examples given in Chapter 5 used a ratio of partial sums of singular values from the reference image. This appeared to give consistent quality for a variety of sequences, but better selection methodologies for block acquisition size may be needed in clinical settings. Furthermore, this number was assumed to be fixed in Chapter 5. It may in fact be advantageous to dynamically set the number of new input vectors for each new image acquisition. Conceivably, this selection would depend on a measured quality of output data, growing larger if the output data is dramatically different than expected and growing smaller if the output data closely matches the expected data. This approach is bolstered by the fact that the theoretically

optimal method can reconstruct high quality images of the synthetic knee sequence using only *one* input vector.

Other secondary issues include how often data needs to be acquired and whether the acquisition and image estimation should be performed a single vector at a time or in a block fashion. Preliminary analysis shows that block-vector acquisitions does not track the image changes quite as well as single-vector acquisitions, but that they are not significantly different. Implementation in a clinical setting may in fact be the driving force in this design decision. Likewise, while most likely image sequence dependent, the rate of data acquisition may depend significantly on hardware constraints. Firm guidelines as to acquisition rates and algorithm robustness for certain clinical sequences needs to be more fully addressed.

Finally, it would be preferable to have a “one size fits all” low-order acquisition method. With the adaptive estimate framework, this would most likely be accomplished with *hybrid* techniques. The primary advantage of the linear system model is that inputs need not be constrained to a single basis. One could potentially use Fourier basis vectors in tandem with vectors identified through other means, e.g., via the SVD, CG-St, or linear predictor methods. This would provide the image estimation with the ability to track both contrast and motion changes while not restricted to one specific modality. Such hybrid techniques are worth further exploration.

# Appendix A

## Analytic Details

### A.1 Linear Algebra Nomenclature

The section seeks to give a brief review of linear algebra terms and concepts required in this thesis. It roughly follows the conventions given by Stewart & Sun in Chapter I of [40].

Throughout this thesis there is a loose association between the letter denoting a vector or matrix and the lower-case letter denoting its elements. Thus,  $a_{ij}$  will usually denote the  $(i, j)$ -element of a matrix  $A$  and  $x_j$  will denote the  $j^{\text{th}}$  element of the vector  $x$ .

The zero vector, or scalar, will all be written  $\mathbf{0}$  (a boldface zero). The zero matrix will be written  $\mathbf{0}$ . The identity matrix will be written as  $I$ , or  $I_n$  when it is necessary to specify the order. The vector, or matrix, of all ones will be written  $\mathbf{1}$  (a boldface one).

The transpose of a matrix  $A$  is denoted  $A^T$  which swaps the rows and columns of the matrix, i.e., the  $(i, j)$ -element of  $A^T$  is  $a_{ji}$ . The complex conjugate transpose of a matrix  $A$  is denoted  $A^H$  which swaps the rows for the conjugate of the columns, that is the  $(i, j)$ -element of  $A^H$  is  $a_{ji}^*$ .

The standard matrix-vector product  $b = Ax$  is defined as  $b_i = \sum_{j=1}^n a_{ij}x_j$ . The element-by-element product, also referred to as the Schur or Hadamard product [18, Chp. 5], is denoted  $C = A \circ B$  where  $c_{ij} = a_{ij}b_{ij}$ .

The matrix  $A$  is

1. SYMMETRIC (HERMITIAN) if  $A^T = A$  ( $A^H = A$ );
2. POSITIVE DEFINITE ( POSITIVE SEMI-DEFINITE, NEGATIVE DEFINITE, NEGATIVE SEMI-DEFINITE) if it is Hermitian and  $x^H A x > (\geq, <, \leq) 0$  for all  $x \neq 0$ ; an equivalent condition on a Hermitian matrix  $A$  is that all of the eigenvalues of  $A$  are positive (non-negative, negative, non-positive);
3. NORMAL if  $A^H A = A A^H$ ;
4. UNITARY, or ORTHOGONAL if  $A^H A = A A^H = I$ ;
5. UPPER TRIANGULAR if it is square and  $\alpha_{ij} = 0 \forall i > j$ ; i.e., if it is zero below its diagonal;
6. LOWER TRIANGULAR if it is square and  $\alpha_{ij} = 0 \forall i < j$ ; i.e., if it is zero above its diagonal;
7. DIAGONAL if it is upper and lower triangular; i.e., its nonzero elements are on its diagonal;
8. a PERMUTATION MATRIX if it is obtained by permuting rows and columns of the identity matrix.
9. IDEMPOTENT if  $A^2 = A$ .
10. SKEW SYMMETRIC (SKEW HERMITIAN) if  $(A^T - A)/2 = A$  ( $(A^H - A)/2 = A$ )

The notation  $\text{diag}\{(\delta_1, \delta_2, \dots, \delta_n)\}$  will mean a DIAGONAL MATRIX whose diagonal elements are  $\delta_1, \delta_2, \dots, \delta_n$ . The scalars  $\delta_i$  may be replaced by square matrices, in which case the matrix will be said to be BLOCK DIAGONAL. BLOCK TRIANGULAR matrices are defined similarly. The notation  $\text{vec}\{A\}$  refers to stacking the columns of matrix  $A$  to create a single vector.

The FROBENIUS NORM of  $A$  is defined as

$$\|A\|_F^2 = \sum_{i,j} |a_{ij}|^2. \quad (\text{A.1})$$

The Frobenius norm may also be calculated as  $\|A\|_F^2 = \text{tr}\{A^H A\} = \sum_i (A^H A)_{ii}$ .

### Projections

Let  $\mathcal{X}$  be a subspace of  $n$ -dimensional Euclidean space and let the columns of  $Q_{\mathcal{X}}$  form an orthonormal basis for  $\mathcal{X}$ . The matrix

$$P_{\mathcal{X}} = Q_{\mathcal{X}} Q_{\mathcal{X}}^H$$

is called the ORTHOGONAL PROJECTION ONTO  $\mathcal{X}$ . Any vector  $z$  can be decomposed into two terms

$$z = P_{\mathcal{X}} z + P_{\mathcal{X}}^{\perp} z$$

where  $P_{\mathcal{X}}^\perp$  is a projection onto the orthogonal complement of  $\mathcal{X}$ .

Projections do not have to be orthogonal. In fact, any idempotent matrix  $P$ , Hermitian or not, can be regarded as an oblique projection onto the range space of  $P$ .

### Eigenvalues and Eigenvectors

The pair  $(x, \lambda)$  is called an EIGENPAIR of the matrix  $A$  if  $x \neq 0$  and  $Ax = \lambda x$ . The set of eigenvalues of  $A$  is written  $\mathcal{L}(A)$ . An eigen vector decomposition of a matrix  $A$  is

$$A = U \text{diag}\{\mathcal{L}(A)\} U^H$$

where the matrix  $U$  is unitary. A unitary matrix is a normal matrix with eigenvalues on the unit circle. A Hermitian matrix is a normal matrix with real eigenvalues.

### The Singular Value Decomposition

A matrix  $A$  can be decomposed into [39]

$$A = U \Sigma V^H \tag{A.2}$$

where  $A$  is a matrix of order  $n$ ,

$$\Sigma = \text{diag}\{(\sigma_1, \sigma_2, \dots, \sigma_n)\}$$

has non-negative elements arranged in decreasing order, and  $U = (u_1 \ u_2 \ \dots \ u_n)$  and  $V = (u_1 \ u_2 \ \dots \ u_n)$  are orthogonal matrices. The elements of  $\Sigma$  are known as the *singular values*, thus A.2 is known as the singular value decomposition (SVD) of  $A$ . The Frobenius norm of  $A$  is related to the singular values via

$$\|A\|_F^2 = \sum_{i,j} |a_{ij}|^2 = \sum_i \sigma_i^2 \tag{A.3}$$

Through the matrix products

$$A^H A = V \Sigma U^H U \Sigma V^H = V \Sigma^2 V^H$$

and

$$A A^H = U \Sigma V^H V \Sigma U^H = U \Sigma^2 U^H$$



we see that  $U$  and  $V$  are the eigen-decomposition unitary matrices of  $AA^H$  and  $A^H A$  respectively.

E. Schmidt showed [39] that if

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^H \quad (\text{A.4})$$

then

$$\|A - A_k\|_F^2 = \|A\|_F^2 - \sum_{i=1}^k \sigma_i^2 = \sum_{i=k+1}^n \sigma_i^2 \quad (\text{A.5})$$

which was subsequently extended to rectangular matrices by Eckart and Young [9, 10]. Thus discarding the least significant singular values and associated vectors provides a method of *matrix approximation*.

The dominant singular vector of  $A$  can be identified [18] through

$$\max_v \|Av\|_F^2.$$

Given that the singular vectors are orthogonal, one could solve the following problem repeatably

$$\max_{v_k} \|(A - A_{k-1})v_k\|_F^2.$$

for  $k = 1, 2, \dots, n$  to identify each singular vector of  $A$  in order.

The ABSOLUTE ERROR of an approximate matrix  $\hat{A}$  is defined as

$$\text{ae}(A, \hat{A}) = \|\hat{A} - A\|_F^2. \quad (\text{A.6})$$

The RELATIVE ERROR in  $\hat{A}$  is

$$\text{re}(A, \hat{A}) = \frac{\|\hat{A} - A\|_F^2}{\|A\|_F^2}. \quad (\text{A.7})$$

The MOORE-PENROSE GENERALIZED INVERSE (also known as PSEUDO-INVERSE) of  $A$  satisfies each of the following conditions:

1.  $AA^\dagger A = A$ ,
2.  $A^\dagger AA^\dagger = A^\dagger$ , and
3.  $AA^\dagger$  and  $A^\dagger A$  are both Hermitian.

By retaining only the  $k$  dominant singular values of  $A$  one can construct the generalized inverse  $A^\dagger$  via

$$A^\dagger = V\Sigma^\dagger U^H = V\text{diag}\{(\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_k^{-1}, 0, \dots, 0)\}U^H \quad (\text{A.8})$$

## A.2 Derivatives of complex valued matrix functions

This section concerns finding the derivative of a complex valued function  $\mathcal{J}(\bar{w})$  with respect to the vector  $\bar{w}$ . The material presented here draws primarily from the discussion given by Haykin in [14, Appendix B].

Given a set of  $N$  complex valued scalar parameters  $w_k = x_k + jy_k$ , one can form a complex valued vector

$$\bar{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix}.$$

We begin by defining the gradient of the function  $\mathcal{J}$  as

$$\nabla \mathcal{J} = 2 \frac{d\mathcal{J}}{d\bar{w}^*}$$

where the derivative and partial derivative operators are defined as

$$\frac{d}{d\bar{w}} \triangleq \frac{1}{2} \begin{bmatrix} \frac{\partial}{\partial x_1} - j \frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} - j \frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_N} - j \frac{\partial}{\partial y_N} \end{bmatrix} \quad \text{and} \quad \frac{\partial}{\partial w_k} = \frac{1}{2} \left( \frac{\partial}{\partial x_k} - j \frac{\partial}{\partial y_k} \right)$$

respectively. Applying the partial derivative to  $w_k$  and the conjugate  $w_k^*$  we find

$$\frac{\partial w_k}{\partial w_k} = \frac{1}{2} \left( \frac{\partial}{\partial x_k} - j \frac{\partial}{\partial y_k} \right) (x_k + jy_k) = 1/2[1 + 0 + 0 - j^2] = 1 \quad (\text{A.9})$$

and

$$\frac{\partial w_k^*}{\partial w_k} = \frac{1}{2} \left( \frac{\partial}{\partial x_k} - j \frac{\partial}{\partial y_k} \right) (x_k - jy_k) = 1/2[1 + 0 + 0 + j^2] = 0 \quad (\text{A.10})$$

Similarly,  $\frac{\partial w_k}{\partial w_k^*} = 0$  and  $\frac{\partial w_k^*}{\partial w_k^*} = 1$ .

Thus, for the following matrix combinations of  $\bar{w}$  and an arbitrary vector  $\bar{p}$  we find:

$$\begin{aligned}\frac{\partial}{\partial \bar{w}}(\bar{w}^H \bar{p}) &= 0 & \frac{\partial}{\partial \bar{w}}(\bar{p}^H \bar{w}) &= \bar{p}^H \\ \frac{\partial}{\partial \bar{w}^*}(\bar{w}^H \bar{p}) &= \bar{p} & \frac{\partial}{\partial \bar{w}^*}(\bar{p}^H \bar{w}) &= 0\end{aligned}$$

### A.3 Efficient solution of vectorized systems

In the development of the CCD algorithm, systems of linear equations appear which are described by matrix equations that contain Kronecker products. Systems containing Kronecker products tend to be very large and require a substantial amount of memory and processing power to solve. However, the system matrices presented in this paper, (4.18) and (4.19), contain a significant level of structure that can be exploited to speed the system solution calculation.

In each of the equations mentioned, the symmetric matrix to invert is of the form  $M^T \text{diag}\{\text{vec}\{S \circ S\}\}M$ . In (4.18),  $M = (L \otimes A) = (L \otimes I_M)(I_r \otimes A^T)$  and likewise in (4.19),  $M = ((AX)^T \otimes I_N)$ . If  $A$  has full row rank, as we assumed throughout this paper, then in each case the central matrix to invert is of the form  $R = (B^T \otimes I) \text{diag}\{\text{vec}\{C\}\}(B \otimes I)$ . For (4.18),  $B = L$ , and for (4.19),  $B = (AX)$ . Expressions of this type can be rewritten in block matrix form with each block containing a diagonal matrix, determined as

$$R_{ij} = (B_{(:,i)}^T \otimes I) \text{diag}\{C\}(B_{(:,j)} \otimes I) = \text{diag}\{C(B_{(:,i)} \circ B_{(:,j)})\}. \quad (\text{A.11})$$

Collecting the sub-blocks together, we find

$$R = \begin{bmatrix} \text{diag}\{C(B_{(:,1)} \circ B_{(:,1)})\} & \text{diag}\{C(B_{(:,1)} \circ B_{(:,2)})\} & \cdots & \text{diag}\{C(B_{(:,1)} \circ B_{(:,m)})\} \\ \text{diag}\{C(B_{(:,2)} \circ B_{(:,1)})\} & \text{diag}\{C(B_{(:,2)} \circ B_{(:,2)})\} & \cdots & \text{diag}\{C(B_{(:,2)} \circ B_{(:,m)})\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{diag}\{C(B_{(:,m)} \circ B_{(:,1)})\} & \text{diag}\{C(B_{(:,m)} \circ B_{(:,2)})\} & \cdots & \text{diag}\{C(B_{(:,m)} \circ B_{(:,m)})\} \end{bmatrix}. \quad (\text{A.12})$$

This matrix can then be permuted to form a block diagonal matrix via

$$P_r R P_c = \begin{bmatrix} \square & & & \\ & \square & & \\ & & \ddots & \\ & & & \square \end{bmatrix}. \quad (\text{A.13})$$

To find the pseudo-inverse of this matrix, one may use the SVD of each individual sub-block to construct the SVD of the entire matrix  $R$ . By decomposing the matrix in this way, the processing resources required to compute the pseudo-inverse can be dramatically reduced. This enables the solutions required in each iteration of the CCD algorithm, § 4.3, to be calculated very quickly.

## A.4 Index of symbols

This thesis presents material from three disciplines: physics, mathematics, and signal processing. While closely related, it should not be a surprise that a symbol associated to a concept in a given discipline will have a completely different association in a different field. There are two conflicting needs when faced with the choice of symbol designations for the concepts presented. On the one hand, the symbols presented should be in close correlation with other published literature. On the other, there are a limited number of symbols, and one would like to avoid symbol conflict — that is having one symbol represent two concepts.

When discussing linear algebra concepts, this work uses capital arabic letters for matrices. There is a loose association between the letter denoting a vector or matrix and the lower-case letter denoting its elements. Thus,  $a_{ij}$  will usually denote the  $(i,j)$ -element of a matrix  $A$  and  $a_j$  will usually denote a row or column of  $A$ . Lower-case greek letters generally represent scalar quantities. Upper-case greek letters represent either the Fourier domain representation of a matrix or a significant scalar function. When discussing physics related material, most vectors represent an element triple in Euclidean space, and are thus represented in **bold** notation. For clarity, the tables below list the symbols used for each chapter.

---

**Chapter 2**


---

$\hat{a}_x, \hat{a}_y, \hat{a}_z$	Euclidean space basis vectors
$\rho$	spin distribution in a volume
$\vec{I}$	intrinsic particle spin vector
$I$	particle spin state
$\vec{\mu}$	molecular magnetic moment
$\gamma$	gyrometric ratio
$\hbar$	Planck's constant divided by $2\pi$
$\mathbf{L}$	angular magnetization vector
$\mathbf{M} = M_x \hat{a}_x + M_y \hat{a}_y + M_z \hat{a}_z$	bulk magnetization vector
$\mathbf{B}_0 = B_0 \hat{a}_z$	strong static magnetic field
$\mathbf{B}_1$	oscillating transverse magnetic field
$\mathbf{G}$	gradient magnetic field
$t$	time
$T$	temperature ( $^{\circ}$ Kelvin)
$E_m$	energy of particle $m$
$\mathbf{r}$	particle position
$\mathbf{v}$	particle velocity
$\omega$	precessional frequency
$\omega_0$	Larmor frequency ( $\omega_0 = \gamma B_0$ )
$\phi$	phase of magnetization vector
$\theta$	tip angle of magnetization vector
$\tau$	time (constant) between successive rf pulses
$T_1$	longitudinal or spin-lattice relaxation time
$T_2$ and $T_2^*$	transverse or spin-spin relaxation time (* indicates recoverable energy)
$\mathbf{k} = k_x \hat{a}_x + k_y \hat{a}_y + k_z \hat{a}_z$	reciprocal spatial distance
$S(t)$ or $S(\mathbf{k})$	received signal (in either time domain or sampled $k$ -space domain)

---

---

**Chapter 3**


---

$S(\mathbf{k})$	received signal
$\rho(\mathbf{r})$	spin distribution at a spatial point $\mathbf{r}$
$\mathbf{k} = k_x \hat{a}_x + k_y \hat{a}_y + k_z \hat{a}_z$	reciprocal spatial distance
$x, y, z$	spatial location
$M, N$	number of matrix rows, columns
$\mathcal{R}$	$k$ -space data matrix from sampled received signal
$c_i$	RIGR model parameters
$u, v$ and $m, n$	RIGR matrix locations
$p$	Linear system model (LSM) input pulse
$T$	LSM duration of phase gradient pulse
$\tau$	LSM total time
$\Delta t_p$	LSM interval between hard pulses
$g_m$	LSM model parameters
$c_m$	LSM fixed set of input pulses with known response
$\Pi$	LSM hard pulse
$\mathcal{R}$	system response matrix
$\mathcal{Y}$	system response in $k$ -space
$\mathcal{P} = Cg$	system input in $k$ -space
$F_N$	Unitary Fourier transform matrix of size $N$
$Y = AX$	linear system model in spatial domain
$\mathcal{E}$	absolute error
$S$	selection matrix
$\theta$	principle angles
$U\Sigma V^H$ and $u, v, \sigma$	SVD components

---



---

**Chapter 4**


---

$\mathcal{J}$	cost
$Y = AX$ and $\hat{A} = YL^T$	linear system model in spatial domain
$S$	selection matrix
$M, N$	matrix sizes
$r$	number of columns in input/output matrices
$Q = XL^T$	special case of reconstruction projection
$\sigma_i$	singular values
$\epsilon$	error threshold
$s_1, s_2$ and $D_1, D_2$	low-rank vectors of $S$ and $\text{diag}\{s_i\}$
$\alpha_i$	logical index vector
$q_i$	columns of $Q$
$i, j$	$(i, j)$ th element of matrix
$s_{ij}$	$(i, j)$ th element of matrix $S$
$Z, W$	arbitrary matrices (mental constructs)

---

---

**Chapter 5**


---

$\mathcal{J}$	cost
$Y_n = A_n X_n$	linear system model in spatial domain
$\hat{A}, \hat{Y}$	estimate of image, output
$M, N$	matrix sizes
$r$	number of columns in input/output matrices
$\alpha, \beta$	generic image change parameters
$\mathcal{E}$	absolute error

---

$\phi(x), A, b$	function for development of CG algorithm
$r_n$	residual
$\alpha$	step parameter
$\beta$	step parameter
$p_k$	CG step direction
$St(n, p)$	Stiefel manifold of order $n$ with $p$ columns
$X, Z$	matrices in $St(n, p)$ definition
$P$	tangential matrix
$QR$	QR decomposition matrices

---

$\tilde{A}_{n+1}$	predicted matrix
$\hat{A}_n$	predictor input matrix
$c$	predictor model coefficients
$u, v, w$	arbitrary coefficients
$RW$	arbitrary matrices
$\mathcal{E}_n =  A - \hat{A} $	absolute error

---



---

**Chapter 6**


---

$S, Y, A, X, L$	static problem matrices
$\mathcal{N}$	Noise matrix

---

# Bibliography

- [1] D. R. Bailes and D. J. Bryant. NMR imaging. *Contemporary Physics*, 25:441, 1984.
- [2] A. Bjorck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, July 1973.
- [3] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*, volume 120 of *Pure and Applied Mathematics*. Academic Press, 2nd edition, 1986.
- [4] M. E. Brummer, W. Dixon, B. Gerety, and H. Tuithof. Composite k-space windows (keyhole techniques) to improve temporal resolution in a dynamic series of images following contrast administration. In *Proc. ISMRM 11th Annual Meeting*, page 4236, 1992.
- [5] P. T. Callaghan. *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford University Press, New York, NY, 1991.
- [6] Y. Cao and D. N. Levin. On the relationship between feature-recognizing MRI and MRI encoded by singular value decomposition. *Magnetic Resonance in Medicine*, 33(1):140, Jan. 1995.
- [7] Y. Cao and D. N. Levin. Using prior knowledge of human anatomy to constrain MR image acquisition and reconstruction: half k-space and full k-space techniques. *Magn Reson Imaging*, 15(6):669–77, 1997.
- [8] N. Cohen, C. R. Johnson, L. Rodman, and H. J. Woerdeman. Ranks of completions of partial matrices. *Operator Theory: Advances and Applications*, 40:165–185, 1989. The Gohberg anniversary collection, Vol. I.
- [9] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.
- [10] C. Eckart and G. Young. A principal axis transformation for non-Hermitian matrices. *Bulletin of the American Mathematical Society*, 45:118–121, 1939.
- [11] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [12] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.
- [13] J. M. Hanson, Z.-P. Liang, R. L. Magin, J. L. Duerk, and P. C. Lauterbur. A comparison of RIGR and SVD dynamic imaging methods. *Magnetic Resonance in Medicine*, 38(1):161–7, July 1997.
- [14] S. Haykin. *Adaptive Filter Theory*. Information and System Sciences. Prentice Hall, 3rd edition, 1996.



- [15] D. M. Healy, Jr and J. B. Weaver. Two applications of wavelet transforms in magnetic resonance imaging. *IEEE Transactions on Information Theory*, 38(2):840–860, Mar. 1992.
- [16] J. Hennel and J. Klinowski. *Fundamentals of Nuclear Magnetic Resonance*. John Wiley & Sons, Inc., 1993.
- [17] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 1985.
- [18] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, NY, 1991.
- [19] A. K. Jain. *Fundamentals of Digital Image Processing*. Information and System Sciences. Prentice Hall, 1989.
- [20] C. R. Johnson, B. Kroschel, and H. Wolkowicz. An interior-point method for approximate positive semidefinite completions. *Comput. Optim. Appl.*, 9(2):175–190, 1998.
- [21] C. R. Johnson and G. T. Whitney. Minimum rank completions. *Linear and Multilinear Algebra*, 28:271–273, 1991.
- [22] F. A. Jolesz and S. M. Blumenfeld. Interventional use of magnetic resonance imaging. *Magn. Reson. Q*, 10(2):85–96, June 1994.
- [23] A. B. Kerr, J. M. Pauly, B. S. Hu, K. C. Li, C. J. Hardy, C. H. Meyer, A. Macovski, and D. G. Nishimura. Real-time interactive MRI on a conventional scanner. *Magnetic Resonance in Medicine*, 38(3):355–67, Sept. 1997.
- [24] P. C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.
- [25] Z.-P. Liang and P. C. Lauterbur. An efficient method for dynamic magnetic resonance imaging. *IEEE Trans. on Medical Imaging*, 13(4):677–686, Dec. 1994.
- [26] Z.-P. Liang and P. C. Lauterbur. *Principles of Magnetic Resonance Imaging : A Signal Processing Perspective*. Biomedical Engineering Series. IEEE, 1999.
- [27] H. L. Liu, Y. Pu, Y. Liu, L. Nickerson, T. Andrews, P. T. Fox, and J. H. Gao. Cerebral blood flow measurement by dynamic contrast MRI using singular value decomposition with an adaptive threshold. *Magnetic Resonance in Medicine*, 42(1):167–72, July 1999.
- [28] D. G. Luenberger. *Linear and Non-Linear Programming*. Addison-Wesley, Reading, MA, 2nd edition, 1984.
- [29] R. Mathias. Matrix completions, norms and Hadamard products. *Proc. Amer. Math. Soc.*, 117(4):905–918, 1993.
- [30] G. A. Morris and R. Freeman. Selective excitation in Fourier transform nuclear magnetic resonance. *Journal of Magnetic Resonance*, 29:433, 1978.
- [31] S. K. Nagle, D. N. Levin, and V. Y. Kuperman. A new sampling theorem for FT MRI of multiple regions. In *Proc. Int. Soc. of Magn. Reson. Med.*, page 187. Magn Reson in Med, 1998.
- [32] W. I. Newman. Extension to the the maximum entropy method. *IEEE Transactions on Image Processing*, IT-23:89, Jan. 1997.

- [33] L. P. Panych, P. D. Jakab, and F. A. Jolesz. An implementation of wavelet encoded MRI. *J. Magn. Reson. Imag.*, 3:649, 1993.
- [34] L. P. Panych, C. Oesterle, G. P. Zientara, and J. Henning. Implementation of a fast gradient-echo SVD encoding technique for dynamic imaging. *J. Magn. Reson. Imag.*, 35:554–562, 1996.
- [35] L. P. Panych, P. Saiviroonporn, G. P. Zientara, and F. A. Jolesz. Implementation of a 3D echo-planar method for SVD encoded MRI. In *Proc. ISMRM 4th Scientific Meeting and Exhibition*, page 387, New York, 1996.
- [36] L. P. Panych, G. P. Zientara, and F. A. Jolesz. MR image encoding by spatially selective rf excitation: An analysis using linear response models. *Int. J. Imaging Syst. Technol.*, 10(2):143–150, 1999.
- [37] D. A. Park. *Introduction to the Quantum Theory*. McGraw-Hill, 2nd edition, 1974.
- [38] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*, 42(5):952–62, Nov. 1999.
- [39] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, pages 551–566, 1993.
- [40] G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.
- [41] P. Strobach. Low-rank adaptive filters. *IEEE Transactions on Signal Processing*, 44(12):2932–2947, Dec. 1996.
- [42] J. Tacke, G. Adam, H. Classen, A. Muhler, A. Prescher, and R. W. Gunther. Dynamic MRI of a hypovascularized liver tumor model. *J. Magn. Reson. Imag.*, 7(4):678–82, Jul-Aug 1997.
- [43] J. van Vaals, M. E. Brummer, and et. al. Keyhole method for accelerating imaging of a contrast agent uptake. *J. Magn. Reson. Imag.*, 3(4):671–675, 1993.
- [44] F. Wallis and F. J. Gilbert. Magnetic resonance imaging in oncology: an overview. *J. R. Coll. Surg. Edinb.*, 44(2):117–25, Apr. 1999.
- [45] G. A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Magazine*, 14(1):56–66, Jan. 1997.
- [46] P. C. Yang, A. B. Kerr, A. C. Liu, D. H. Liang, C. Hardy, C. H. Meyer, A. Macovski, J. M. Pauly, and B. S. Hu. New real-time interactive cardiac magnetic resonance imaging system complements echocardiography. *J. Am. Coll. Cardiol.*, 32(7):2049–56, Dec. 1998.
- [47] G. P. Zientara, L. P. Panych, and F. A. Jolesz. Dynamically adaptive MRI with encoding by singular value decomposition. *Magnetic Resonance in Medicine*, 32:268–274, 1994.