SPARSE RECONSTRUCTION METHODS IN FLUORESCENCE DIFFUSE OPTICAL TOMOGRAPHY

By

An Jin

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
Major Subject: BIOMEDICAL ENGINEERING

Approved by the
Examining Committee:

Birsen Yazıcı, Thesis Adviser

George Xu, Member

Suvranu De, Member

Partha Dutta, Member

Rensselaer Polytechnic Institute
Troy, New York

November 2012
(For Graduation December 2012)

CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

ACKNOWLEDGMENT xiii

ABSTRACT xiv

1. INTRODUCTION 1
    1.1 Diffuse Optical Tomography and Fluorescence Diffuse Optical Tomography 1
    1.2 FDOT Imaging Problem 2
        1.2.1 FDOT Forward Problem 2
        1.2.2 Nonlinear and Linear FDOT Inverse Problems 4
        1.2.3 Challenges in FDOT Inverse Problem 4
    1.3 The Objective of the Thesis 4
    1.4 Regularization Methods for FDOT Image Reconstruction 5
    1.5 Compressive Sensing Framework for FDOT Image Reconstruction . . 5
        1.5.1 Overview in Compressive Sensing? 5
        1.5.2 The Sparseness of Fluorophore in FDOT Images 6
        1.5.3 Sparse Signal Recovery Techniques 6
        1.5.4 Performance Guarantees in Compressive Sensing 7
    1.6 Related Literature - Applications of CS in FDOT and DOT 8
    1.7 Contributions of the Thesis 9
    1.8 Thesis Outline 10

2. FDOT IMAGING PROBLEM 12
    2.1 Introduction 12
    2.2 Light Propagation Model for FDOT 15
    2.3 FDOT Forward and Inverse Problems 18
        2.3.1 Measurement Model 18
        2.3.2 Linearization of the FDOT Imaging Problem 18
            2.3.2.1 Weak Fluorophore Linear Approximation 18
            2.3.2.2 Nonlinear Iterative Perturbation 19

2.3.3 Forward Operator Formulation . . . 20
2.4 Discretization of the FDOT Imaging Problem . . . 20
2.4.1 Forward Problem Discretization . . . 20
2.4.2 Inverse Problem Discretization . . . 24
2.5 Linear FDOT Statistical Model with Additive Noise . . . 26
2.6 Conclusion . . . 27
3. REGULARIZATION METHODS IN FDOT . . . 28
3.1 Introduction . . . 28
3.2 Linear Regularization Techniques in Literature . . . 29
3.2.1 Truncated Singular Value Decomposition . . . 29
3.2.2 Conjugate Gradient Method . . . 31
3.2.3 Kaczmarz Iteration Method . . . 32
3.2.4 Tikhonov Regularization Method . . . 33
3.2.5 The L_p-Norm ($1 \leq p < 2$) Regularization . . . 34
3.3 The Combined L_2-L_p-Norm Regularization . . . 36
3.3.1 *A Priori* Information in FDOT . . . 36
3.3.2 The Combined L_2-l_p-Norm Regularization with Specific *A Priori* Information . . . 37
3.4 Numerical Simulations . . . 39
3.4.1 Simulation Setup . . . 40
3.4.2 Singular Value Spectrum of the Forward Matrix . . . 41
3.4.3 Visual Reconstruction Results . . . 42
3.4.3.1 Phantom with Single Heterogeneity . . . 42
3.4.3.2 Phantom with Two Heterogeneities . . . 43
3.4.4 Quantitative Performance Measurements . . . 46
3.5 Silicon Phantom Experiment . . . 52
3.6 Conclusion . . . 53
4. SPARSE SIGNAL RECOVERY TECHNIQUES IN FDOT RECONSTRUCTION . . . 63
4.1 Introduction . . . 63
4.2 Sparse Representation . . . 64
4.3 Signal Recovery with Sparsity Constraint . . . 65
4.4 Incoherent Sampling . . . 65

4.5 Sparse Signal Recovery Techniques in CS Literature 67
    4.5.1 Greedy Algorithms 68
        4.5.1.1 Orthogonal Matching Pursuit 68
        4.5.1.2 Gradient Pursuit 69
        4.5.1.3 Stagewise Orthogonal Matching Pursuit 70
        4.5.1.4 Regularized Orthogonal Matching Pursuit 72
        4.5.1.5 Acrostic Compressive Sampling Matching Pursuit . . 73
    4.5.2 Convex Relaxation 74
        4.5.2.1 BP-simplex 75
        4.5.2.2 BP-interior 75
        4.5.2.3 Iterative Shrinkage/Thresholding 77
        4.5.2.4 Gradient Projection for Sparse Reconstruction 79
        4.5.2.5 Focal Underdetermined System Solver 80
    4.5.3 Non-convex Relaxation 81
4.6 Numerical Simulations 82
    4.6.1 Simulation Setup 82
    4.6.2 Imaging Reconstruction - Visual Results 83
    4.6.3 Imaging Reconstruction - Quantitative Measurements 85
4.7 Silicon Phantom Experiment 86
4.8 Discussion 87
4.9 Conclusion 88

5. FDOT RECONSTRUCTION WITH PRECONDITIONED FORWARD SENSING MATRIX 94
5.1 Introduction 94
5.2 Preconditioned FDOT Inverse Problem with Sparsity Constraint . . . 95
5.3 Performance Guarantees in Sparse Signal Reconstruction 96
5.4 Preconditioning Matrix Design 97
    5.4.1 Elad's Method 97
    5.4.2 Duarte-Carvajalino and Sapiro's Method 98
    5.4.3 Schnass and Vandergheynst's Method 99
    5.4.4 Our Approach 99
5.5 Numerical Simulations 100
    5.5.1 Simulation Setup 101
    5.5.2 Simulation Results - Coherence of the FDOT Forward Matrix 102

5.5.3 Simulation Results - Image Reconstruction 103
5.6 Silicon Phantom Experiment 106
5.6.1 Real Phantom Configuration 106
5.6.2 Coherence of the FDOT Forward Matrix 108
5.7 Conclusion 109
6. LIGHT ILLUMINATION AND DETECTION PATTERNS FOR FDOT BASED ON COMPRESSIVE SENSING 113
6.1 Introduction 113
6.2 Bounds on the Coherence of the FDOT Forward Matrix 114
6.3 The FDOT Forward Matrix Optimization 118
6.3.1 Design of the Illumination Patterns - The Optical Mask 119
6.3.2 Design of Detection Patterns - The Measurement Mask 122
6.3.3 Forward Sensing Matrix Construction 124
6.4 Numerical Simulations 125
6.4.1 Simulation Setup 125
6.4.2 Simulation Results - Coherence of the FDOT Forward Matrix 127
6.4.3 Simulation Results - Image Reconstruction 128
6.5 Silicon Phantom Experiment 130
6.5.1 Real Phantom Configuration 131
6.5.2 Coherence of the FDOT Forward Matrix 131
6.5.3 Image Reconstruction 132
6.6 Conclusion 132
7. CONCLUSION AND FUTURE WORK 141
7.1 Conclusion 141
7.2 Future Work 143
LITERATURE CITED 145

APPENDICES

A. Calculation of the Combined l_2-l_p norm Regularization Coefficients 161
B. Upper Bounds for Cumulative Coherence 163
C. Equivalence of the Normalized Forward Sensing Matrix 166

LIST OF TABLES

2.1 Table of notations 12

3.1 Spatial resolution of the reconstructed images, $r = 0.5$ cm, $d = 1.0$ cm. . 51

3.2 Spatial resolution of the reconstructed images, $r = 0.5$ cm, $d = 1.5$ cm. . 52

3.3 Spatial resolution of the reconstructed images, $r = 0.5$ cm, $d = 2.0$ cm. . 52

3.4 The CONR and CBNR for different regularization methods. 53

4.1 NMSE of the reconstructed images (10^{-4}) using different sparse signal recovery techniques. 85

4.2 PVR of the reconstructed images using different sparse signal recovery techniques. 86

5.1 NMSE of the reconstructed images (10^{-4}). 106

5.2 CBNR of the reconstructed fluorophore images using different algorithms. 109

6.1 NMSE of the reconstructed images (10^{-4}). 130

LIST OF FIGURES

3.1 A comparison of the cost functions of the Tikhonov regularization and l_p-norm regularization. 35

3.2 A comparison of the cost functions of the Tikhonov regularization and the combined l_2-l_p-norm ($p = 1$) regularization. 38

3.3 Digital phantom simulation setup. 40

3.4 Singular value spectrum of the forward matrix. 41

3.5 Fluorophore configuration of the Phantom with single heterogeneity of radius 0.5 cm. 42

3.6 The SVD reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 43

3.7 The CG reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 43

3.8 The ART reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 44

3.9 The Tikhonov reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 44

3.10 The l_1-norm reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 45

3.11 The l_p-norm ($p = 1.2$) reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 45

3.12 The combined l_2-l_p-norm ($p = 1$) reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. . 46

3.13 Fluorophore configuration of phantom with two heterogeneities of radius 0.5 cm (2 cm apart in the center). 46

3.14 The SVD reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. . . 47

3.15 The CG reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. . . 47

3.16 The ART reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. . . 48

3.17 The Tikhonov reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 48

3.18 The l_1-norm reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 49

3.19 The l_p-norm ($p = 1.2$) reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 49

3.20 The combined l_2-l_p-norm ($p = 1$) reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 50

3.21 MSE-SNR plot of six different phantom configurations. 55

3.22 CONR-SNR plot of six different phantom configurations. 56

3.23 CONR-SNR plot of six different phantom 6 configurations. 57

3.24 CE-SNR plot of different phantom configurations. 58

3.25 Examples of the resolutions of heterogeneities. 59

3.26 Fluorophore configuration of the real silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D surface of the heterogeneity. 59

3.27 SVD reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 60

3.28 CG reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 60

3.29 ART reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 60

3.30 The Tikhonov reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 61

3.31 The l_1-norm reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 61

3.32 The l_p ($p = 1.2$) norm reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 61

3.33 The combined $l_2 - l_p$-norm reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction. 62

4.1 Simulation setup. 82

4.2 The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$ cm). 83

4.3 The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using different algorithms, $r = 0.5$ cm, 1% noise. 89

4.4 The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$ cm, $d = 1.5$ cm). 90

4.5 The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using different algorithms, $r = 0.5$ cm, $d = 1.5$ cm, 1% noise. . . 91

4.6 The configuration of the real silicon phantom and the cross section of the fluorophore yield. 92

4.7 The imaging system used in the silicon phantom experiment. 92

4.8 The cross sections of the phantom at $z = 1$ (middle) using different algorithms. 93

5.1 An illustration of the phantom and source-detector configuration. 101

5.2 The normalized inner products and cumulative coherence of the forward matrix before and after applying the preconditioning matrix. 102

5.3 The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$ cm). 103

5.4 The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using greedy algorithms, $r = 0.5$ cm, 1% noise. 104

5.5 The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using convex relaxation algorithms, $r = 0.5$ cm, 1% noise. . . . 105

5.6 The configuration of the real silicon phantom and the cross section of the fluorophore yield. 107

5.7 The imaging system used in the silicon phantom experiment. 107

5.8 The normalized inner products and cumulative coherence of the forward matrix before and after applying the preconditioning matrix. 110

5.9 The cross sections at $z = 1$ (middle) of the reconstructed phantom using greedy algorithms. 111

5.10 The cross sections at $z = 1$ (middle) of the phantom using convex relaxation techniques. 112

6.1 The block diagram of the FDOT imaging process with optical and measurement masks. 124

6.2 An illustration of the phantom and source-detector configuration. 126

6.3 The optical and measurement masks used with simulated data. 126

6.4 The normalized inner products and cumulative coherence of the forward matrix before and after applying optical and measurement masks. . . . 127

6.5 The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$cm). 128

6.6 The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using greedy algorithms, $r = 0.5$ cm, 1% noise. 134

6.7 The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using convex relaxation algorithms, $r = 0.5$ cm, 1% noise. . . . 135

6.8 The configuration of the real silicon phantom and the cross section of the fluorophore yield. 136

6.9 The imaging system used in the silicon phantom experiment. 136

6.10 The optical and measurement masks used with simulated data. 137

6.11 The normalized inner products and cumulative coherence of the forward matrix before and after applying optical and measurement masks. . . . 138

6.12 The cross sections of the reconstructed phantom at $z = 1$ (middle) using greedy algorithms. 139

6.13 The cross sections of the phantom at $z = 1$ (middle) using convex relaxation techniques. 140

ACKNOWLEDGMENT

First of all, I have to thank my parents for their unconditional love and support over all these years. I love them to death, and this Ph.D thesis is dedicated to them. There is always a lot more than I can say in words.

I would like to specially thank my advisor Dr. Birsen Yazici for inviting me to her research program. She is a great professor, and a great woman. There are a number of ups and downs through my Ph.D years, but she has always been patient and supportive. She taught me so much, helped me to learn, and to mature. Dr. Yazici, thank you for investing your precious time in me to make me smarter and stronger. I love you a lot.

I would like to thank my other committee members, Prof. George Xu, Prof. Partha Dutta, and Prof. Suvranu De for dedicating their precious time on my thesis work. I appreciate their indispensable advice to my thesis work.

I would like to thank my former colleague, Dr. Lu Zhou, and my current colleague, Zhengmin Li. Thank you for coming to my life being great colleagues and great friends. It is my pleasure to know you and work with you.

Finally, thanks to all my friends and my families for always encouraging me and cheering me up. You are in my heart no matter where I am. I'm proud to have you all in my life.

ABSTRACT

Fluorescence diffuse optical tomography (FDOT) is an imaging modality that uses near infrared (NIR) light to excite the fluorophore injected into the tissue, and to reconstruct the fluorophore concentration from boundary measurements. FDOT image reconstruction is a highly ill-posed inverse problem. However, the fluorophore distribution in FDOT images is often very sparse. This thesis aims to develop methods and algorithms that exploit the sparsity of fluorophore distribution to address the ill-posedness of the FDOT image reconstruction problem.

The first part of the thesis focuses on designing sparsity promoting regularization combined with specific *a priori* information for FDOT image reconstruction. The second part of the thesis focuses on addressing the ill-posedness of the FDOT inverse problem in the compressive sensing (CS) framework. CS theory guarantees exact recovery of sparse signals if the underlying forward sensing matrix is incoherent. We introduce two methods to reduce the coherence of the FDOT forward sensing matrix, and thereby to improve the image reconstruction performance: (1) We design a preconditioning matrix applied to the FDOT forward sensing matrix at the image reconstruction stage. (2) We design CS-based light illumination and detection patterns that can be applied at the data acquisition stage. We demonstrate the performance of our methods in 3D numerical simulations, and using real data obtained from a phantom experiment.

Finally, we note that while our interest lies primarily in the FDOT imaging problem, the methods and algorithms developed in this thesis are also applicable to other partial differential equation-based inverse coefficient estimation problems.

CHAPTER 1
INTRODUCTION

1.1 Diffuse Optical Tomography and Fluorescence Diffuse Optical Tomography

Diffuse optical tomography (DOT) is an imaging modality that uses light in the *near infrared* (NIR) spectrum to provide volumetric views of optical absorption and scattering coefficients [1–3]. Light in the NIR spectrum ($700 - 900$ nm) can propagate through biological tissue for several centimeters before it is totally attenuated. This is due to the significantly lower optical absorption [4] of NIR light in tissue as compared to the light in the visible range of the spectrum. The NIR light that propagates through the tissue is either elastically scattered or absorbed by water, oxygenated hemoglobin (HbO), and deoxygenated hemoglobin (Hb). The light-tissue interactions, i.e., multiphoton absorption and scattering, corresponds to photophysical processes at the molecular level. In diseased tissue (e.g., tumor), the scattering and absorption coefficients are larger than those in the normal tissue. Thus, DOT offers an opportunity to determine the hemoglobin concentration and the oxygen saturation in tissue, which can be used for screening or diagnosis of diseased tissue. However, in DOT, detection and diagnosis of the diseased tissue are based only on the endogenous chromophores, which does not always provide sufficient contrast.

Fluorescence diffuse optical tomography (FDOT) is an imaging technique for small animals or deep tissue based on similar problems as that of DOT. Unlike DOT, FDOT uses exogenous contrast agents, which are often called fluorescence probes or fluorophores, to increase the optical contrast [5–8]. These fluorophores are molecules [9, 10] that can absorb the energy from an external light source and re-emit light at a lower-energy level at longer wavelengths. In recent years, due to the increasing availability of fluorescent contrast agents [11, 12], the use of fluorophores as exogenous contrast agents for imaging has shown great promise [13–20]. There is a wide application of FDOT in the clinical world, such as drug discovery, tumor detector, etc. The fluorophore probes play a significant role in identifying specific molecules in drug or diseased tissue that would be otherwise impossible to detect. 3D imaging of fluorophore concentration map is able to indicate optical heterogeneity in biological tissue. In experimental studies with animal subjects [13–17, 19], fluorophores have been successfully used to visualize cancerous tissue in vivo near the skin surface. In addition, Indo-Cyanin Green (ICG) enhanced optical tomography has been used to image the absorption of a malignant tumor in human tissue [17,20].

As compared to the traditional radiological imaging modalities, such as X-ray Computer Tomography (X-ray CT) and Magnetic Resonance Imaging (MRI), FDOT offers several advantages: (1) NIR radiation is minimally invasive, thus can be safely used for screening and continuous monitoring. Some fluorescence contrast agents, such as ICG, were experimentally shown to be clinically feasible [21]; (2) the instrumentation for FDOT is relatively inexpensive and portable; (3) FDOT offers the capability to study biological function at the molecular level [22].

1.2 FDOT Imaging Problem

In FDOT, NIR light at the fluorophore's excitation wavelength is directed into a bounded imaging domain to excite the injected fluorophores inside the domain, which, in turn, emits light at a longer (emission) wavelength. The FDOT imaging problem has two arts: the forward and inverse problems. The FDOT forward problem estimates the photon density field at the excitation and emission wavelengths given the optical properties, which include fluorophore concentration, diffusion and absorption coefficients, of the imaging domain. The FDOT inverse problem, on the other hand, uses the excitation and emission photon density field measured at the boundary of the imaging domain to estimate the optical properties of the medium.

1.2.1 FDOT Forward Problem

The forward problem of FDOT involves solving for the light field in the imaging domain based on a mathematical model of light propagation in biological tissue. The radiation transport equation (RTE) and the diffusion equation are the two main models commonly used for photon propagation in highly scattering media. RTE models the propagation of the multiply scattered light, especially in the vicinity of the boundary or where the medium properties vary rapidly in a accurate way. However, the high computational complexity of solving RTE makes it unsuitable for many problems of practical interest [2]. The diffusion equation is an approximation to the RTE, when the scattering coefficients are much larger than the absorption coefficients in the optical domain of interest. It has been shown that there are many situations where diffusion equation can be used to accurately describe the photon migration in highly scattering media [23–25].

Depending on the specific FDOT applications, there are three types of approaches to solve the FDOT forward problem, including analytical, statistical, and numerical methods:

1. Analytical approach attempts to obtain a closed form solution to the diffusion equations. However, analytical solutions only exist for simple imaging domain geometries and homogeneous optical properties [26, 27]. In general, there is no closed form relationship between the optical coefficients and the photon density field for arbitrary imaging domain geometries.

2. Statistical approach attempts to solve the forward problem based on the RTE by simulating a large number of photons. One of the most widely used methods is the Monte Carlo method [28]. Statistical methods are computationally intensive. Therefore, they are typically only used for problems where diffusion equation does not hold [28, 29].

3. Numerical approach employs numerical methods, such as *finite element method* (FEM), to solve the partial differential equations involved in FDOT forward model. This approach has the advantages of computational efficiency as compared to statistical methods and the applicability of arbitrary imaging domain boundaries such as the ones in many practical applications [30–32].

In this thesis, we use two coupled diffusion equations to model the propagation of the excitation and emission light in the optical domain, and use FEM to solve the FDOT forward problem [33, 34, 34–36].

1.2.2 Nonlinear and Linear FDOT Inverse Problems

In the FDOT inverse problem, the objective is to reconstruct the spatially varying optical properties of a medium of interest from boundary measurements. In this thesis, we are primarily interested in reconstructing the fluorophore yield, which is a quantity proportional to the fluorophore concentration.

In FDOT, the fluorophore yield is nonlinearly related to the measurement data. The nonlinear forward operator depends on the absorption coefficients of the imaging domain, which is composed of the endogenous absorption of the tissue and the exogenous absorption of the fluorophore. For the case of weak fluorophore, the contribution of the fluorophore absorption to the total absorption is assumed to be small. Thus, under the assumption that the absorption coefficient is approximately equal to the the endogenous absorption coefficient, the FDOT inverse problem can be linearized. For the case of strong fluorophore, where the fluorophore absorption is not negligible as compared to the total absorption, an iterative linearization approach can be used to address the FDOT inverse problem.

1.2.3 Challenges in FDOT Inverse Problem

In FDOT, the number of measurements available is often much smaller than the number of unknowns, which results in the non-uniqueness of the solution. Additionally, the solution is not stable, i.e., a small perturbation in the measurement data can lead to large deviations in the reconstructed image. Furthermore, FDOT image reconstruction is inherently a three dimensional (3D) problem with a large number of unknowns. Thus, the FDOT inverse problem is an ill-posed and computationally expensive image reconstruction problem.

1.3 The Objective of the Thesis

The overarching goal of this thesis is to address the ill-posedness of the FDOT inverse problem. Towards this goal, we pursue two approaches:

1. Developing a novel regularization method that suits to the specific nature of the FDOT image reconstruction problem within generalized Tikhonov regularization framework.

2. Exploiting the sparsity of the fluorophore yield in the image domain and addressing the ill-posedness of the FDOT inverse problem within the *compressive sensing* (CS) framework.

Below we explain these two approaches in more details.

1.4 Regularization Methods for FDOT Image Reconstruction

Regularization techniques are commonly used to address the ill-posedness of the DOT or FDOT inverse problems. These include the *truncated singular value decomposition* (TSVD) method [37], the *conjugate gradient* (CG) method [38], the Kaczmarz iteration method [39], the Tikhonov regularization method [40–42], the l_p-norm regularization method [43], and others [44].

We develop a combined l_2-l_p-norm regularization technique to address the ill-posedness of the FDOT inverse problem. The new method has the advantage of preserving the signal in the foreground region from over-smoothing while effectively removing the background noise. We demonstrate the performance of the new regulation scheme in 3D numerical simulations and real silicon phantom experiment.

1.5 Compressive Sensing Framework for FDOT Image Reconstruction

1.5.1 Overview in Compressive Sensing?

Recently, the emerging field of compressive sensing has shown that sparse signals, i.e, signals that have a small number of non-zero entries, can be recovered from far fewer samples or measurements than required by the traditional Shannon-Nyquist sampling theorem. The theory of CS was first introduced in [45]. This study considered signals that have a sparse representation in the Fourier domain. In subsequent studies [46], the CS theory was generalized to recover sparse signals in general bases from limited number of measurement data. In nature, a wide range of signals can be sparsely represented in appropriate basis spaces. Therefore, CS theory has been widely applied in many different fields, including compressive imaging [47], medical imaging [48], radar [49], and computational biology [50].

There are three fundamental aspects of CS theory:

1. Signals to be reconstructed must be sparse.
2. Certain optimization methods, which we refer to as the *sparse signal recovery techniques*, can be used to determine sparse signals from an under determined system of linear equations.
3. Under certain conditions on the forward operator, the CS theory provides performance guarantees for the exact recovery of the sparse signals (even in the presence of measurement noise). These conditions are given in terms of the incoherence of the sensing matrix that maps the unknown signal of interest to measurements [51–55].

1.5.2 The Sparseness of Fluorophore in FDOT Images

The fluorophores used in FDOT are typically designed to accumulate in certain regions of interest, i.e., cancerous tumors. The fluorophore concentration is often localized in relatively small, specific regions as compared to the entire image domain. For example, early stage tumors tagged with fluorophore are often very small in size [56]. In some applications, it may be sufficient to model fluorophore concentration as a single heterogeneous point in the entire imaging domain [57–59]. For example, [59] employs an FDOT system with reflection geometry verifies that a point-heterogeneity model is sufficient to represent tumors smaller than 1 cm^3 in size. Thus, in many practical applications, it is reasonable to assume that the fluorophore concentration map to be reconstructed is often very sparse in the image domain. This observation motivates us to address the FDOT image reconstruction problem as a sparse signal recovery problem within the CS framework.

1.5.3 Sparse Signal Recovery Techniques

Commonly used sparse signal recovery strategies are based on greedy-type search strategies and relaxation techniques. The greedy approach has many popular strategies including *orthogonal matching pursuit* (OMP) [60], *gradient pursuit* [61], *stagewise orthogonal matching pursuit* (StOMP) [62], *regularized orthogonal matching pursuit* (ROMP) [63, 64], *acrostic compressive sampling matching pursuit* (CoSaMP) [65], and others [66]. Another popular approach is based on convex relaxation, which often involves solving an l_1-norm constraint optimization problem. A large number of algorithms have been developed to solve the relaxed convex problem, including the linear programming techniques used in *basis pursuit* (BP) [67], the *iterative shrinkage/thresholding* method [68–70], the *gradient projection for sparse reconstruction* (GPSR) method [71], the *focal underdetermined system solver* (FOCUSS) [72, 73], and others [66]. Besides convex relaxation, the non-convex relaxation is also used as a sparsity constraint. A typical non-convex relaxation is to replace the l_1-norm constraint in convex relaxation with the l_p-norm ($0 < p < 1$) constraint. It has been shown that, as compared to the l_1-norm constraint optimization, l_p-norm constraint optimization can recover sparse signals with fewer measurements [74].

1.5.4 Performance Guarantees in Compressive Sensing

An important aspect of the CS theory is that it provides performance guarantees for accurate or exact recovery of sparse signals under some conditions. One of these conditions is the *incoherency* of the so called *sensing matrix* or the forward operator [51, 75, 76]. Intuitively speaking, incoherency is a degree of the orthogonality of the forward matrix. In general, CS theory states that if the coherence of the forward matrix is low enough, a sparse signal can be exactly recovered. However, NIR light is diffusive in nature, thus the excitation or emission light fields at adjacent locations in the imaging domain is similar due to light scattering. As a result, the columns of the FDOT forward sensing matrix are usually coherent. However, in many applications, the forward matrix can be designed to be incoherent, which improves the recovery performance of sparse signals [52–54, 66, 77, 78].

The nature of the forward matrix in FDOT depends on (1) the geometry of the imaging domain; (2) the endogenous optical properties of the medium; (3) the source and detector configurations. The geometry and the endogenous optical properties are inherent properties of the imaging domain, which cannot be changed. However, it is possible to improve the incoherence of the FDOT forward matrix by optimizing the light illumination and detection patterns. In this thesis, we design light illumination and detection patterns for FDOT based on the CS theory to improve the quality of reconstructed images.

1.6 Related Literature - Applications of CS in FDOT and DOT

In FDOT, the imaging domain is often illuminated sequentially by a number of point sources placed at the boundary of the imaging domain. However, recent work has demonstrated that the "structured illumination" can be useful in reducing the ill-posedness of the optical tomography inverse problem, and to improve the spatial resolution of the reconstructed images [79–86]. In [79], a method is presented which designs distribution of light intensities over a predetermined set of locations and modulation frequencies to maximize the detectability of heterogeneities. In [80], different illumination light patterns including line sources, Gaussian spots, and their combination are explored to reduce the ill-posedness of the FDOT problem. In [84] and [85], spatially modulated frequency diversity is explored to improve image reconstruction. In [86], the superiority of wide-field patterned light sources over point sources is demonstrated using phantom experiments. Note that the concept of structured illumination is similar to the coded aperture imaging technique, where predetermined temporal or spatial masks/patterns are used to encode the illumination sources to optimize imaging processes, including the ones in medical imaging [87–89], high energy astronomy [90], spectral imaging [91, 92], and others [93, 94].

Recently, CS framework has been utilized to address the inverse problems of optical tomography [48, 95–98]. In [48], sparse fluorophore concentration is reconstructed by solving an l_1-norm constraint optimization problem. Experimental results obtained from a milk-based phantom demonstrate that the sparsity promoting signal recovery technique is able to accurately localize a small fluorophore heterogeneity in deep phantom. In [95], improvements in DOT image reconstruction is demonstrated in numerical simulations using a greedy-type algorithm in conjugate with a small number of measurements. In [96, 97], the advantages of the CS based methods over the traditional singular value decomposition (SVD) method are numerically demonstrated in DOT. In [98], a non-iterative inversion method based on generalized MUSIC algorithm is introduced, and improvements in DOT image reconstruction are shown in 3D numerical simulations.

Although sparse signal recovery techniques have shown promising results in optical tomography, the existing studies in this field is limited in scope and depth. CS theory has not been explored to improve the incoherence of the FDOT forward sensing matrix.

1.7 Contributions of the Thesis

The contribution of the thesis can be summarized as follows:

1. We evaluate the performance of existing regularization techniques and propose a new one that combines specific *a priori* information with sparsity promoting constraint on fluorophore concentration:
    - We evaluate the existing state-of-art regularization techniques for the FDOT image reconstruction problem, and extensively compare their performances in 3D numerical simulations and in a real silicon phantom experiment [99].
    - We design a novel regularization scheme, which we refer to as the combined l_2-l_p-norm regularization, for the FDOT image reconstruction that takes into account specific *a priori* information available about the fluorophore distribution in tissue while promoting sparsity.
2. We exploit the sparse nature of the fluorophore concentration map in the imaging domain, and evaluate the performance of different sparse signal recovery techniques, including greedy type algorithms and relaxation techniques, in CS literature in FDOT image reconstruction: We show that, in general, the sparsity promoting algorithms provide better reconstruction results than the traditional regularization methods without the sparsity constraint [100].

3. Based on CS framework, we design a preconditioning matrix to improve the incoherence of the FDOT forward matrix, and address the preconditioned FDOT inverse problem using different sparse signal recovery methods [101].
4. We design light illumination and detection patterns to improve the incoherence of the FDOT forward matrix, and hence, improve the FDOT reconstruction performances [102].
5. We tested the performance of these methods extensively in 3D numerical simulation and real data obtained from a phantom experiment.

1.8 Thesis Outline

The rest of the thesis is organized as follows:

In Chapter 2, we present the two coupled diffusion equations to model the NIR light propagation at the excitation and emission wavelengths in a bounded optical domain. Based on the assumption of weak fluorophore, we linearize the FDOT imaging problem by assuming that the contribution of the fluorophore absorption is negligible compared to the total absorption. Next, with an appropriate boundary condition, we present the discretized FDOT imaging problem using FEM with first order Lagrange basis functions; and define the FDOT forward sensing matrix.

In Chapter 3, based on the linearized formulation of the FDOT imaging problem derived in Chapter 2, we apply a number of regularization techniques to address the ill-posedness of the FDOT inverse problem. In addition, we propose a novel regularization technique based on a combined l_2-l_p-norm regularization function.

In Chapter 4, we address the linearized FDOT inverse problem by exploiting the sparse nature of the fluorophore yield in the imaging domain. The sparsity constraints are introduced into the FDOT inverse problem formulation. We have briefly reviewed a number of sparsity promoting reconstruction techniques in CS literature and applied these techniques to the reconstruction of the fluorophore concentration map.

In Chapter 5, in addition to using sparse signal recovery techniques as discussed in Chapter 4, we improve the incoherence of the FDOT forward sensing matrix by applying a preconditioning matrix. We solve the preconditioned FDOT imaging problem, and show the improvements in image reconstruction when the preconditioning matrix is applied.

In Chapter 6, we improve the incoherence of the FDOT forward sensing matrix by designing light illumination and detection patterns. Unlike in Chapter 5, where the FDOT forward sensing matrix is preconditioned directly, we take an alternative approach by decomposing the FDOT forward matrix into the Kronecker product of two underlying matrices, and design two preconditioners to reduce the coherence of these two matrices. Practical implementation of this preconditioning approach results in adjusting the intensities of light sources operating simultaneously, and filtering each set of measurements prior to reconstruction.

Finally, in Chapter 7, we conclude the thesis and provide a discussion on the potential future work.

CHAPTER 2
FDOT IMAGING PROBLEM

2.1 Introduction

In this chapter, we present the FDOT imaging problem formulation. We focus on the estimation of the fluorophore concentration map in a bounded optical domain using the continuous wave method. For the forward problem, the light propagation in this domain is modeled by a pair of coupled frequency-domain diffusion equations with the Robin-type boundary conditions. In general, the FDOT imaging problem is nonlinear. However, based on the assumption of weak fluorophore, the FDOT imaging problem can be linearized by assuming that the contribution of the fluorophore absorption is negligible compared to the total absorption. When strong fluorophore is present in the imaging domain, the FDOT imaging problem can be linearized by applying an iterative perturbation approach. In this thesis, our calculation is based on the linearization approach. We use the FEM and first order lagrange basis function to discretize the solve the FDOT imaging problem.

We use the following notational conventions throughout the thesis. The bold symbol, bold italic lower-case letters, $\boldsymbol{\Gamma}, \boldsymbol{y}$, etc. are used to denote vector variables, italic symbol or Roman letters are used to denote scalar variables. The calligraphic letters, $\mathcal{A}, \mathcal{B}$, etc. are used to denote operators. The bold Roman upper-case letters are used to denote matrices, $\boldsymbol{A}, \boldsymbol{B}$, etc. We use $E_{x,m}$ as a shorthand notation for the quantity E at either excitation or emission wavelengths. Table 2.1 provides a summary of the key variables and notations used throughout the thesis.

Table 2.1: Table of notations

Symbol	Designation
Ω	Bounded imaging domain

continued...

continued from previous page

Symbol	Designation
ω	Angular modulation frequency
N	Number of voxels in the imaging domain
N_s	Number of light sources
N_d	Number of detectors
M	Number of measurements
s^i	i^{th} point source
$\phi^i_{x,m}$	Photon density field due to the i^{th} source
$\tilde{\phi}^i_{x,m}$	Finite element approximation of $\phi^i_{x,m}$
$g^j_{x,m}$	Green's functions corresponding to the j^{th} detector
$\tilde{g}^j_{x,m}$	Finite element approximation of $g^j_{x,m}$
$D_{x,m}$	Diffusion coefficients
$\mu_{ax,m}$	Absorption coefficients
$\mu_{aex,m}$	Endogenous absorption coefficients
$\mu_{afx,m}$	Fluorophore absorption coefficients
μ	Fluorophore yield
$\tilde{\mu}$	Finite element approximation of μ
$\mathbf{\Gamma}$	Vector of measurements
$\mathcal{B}^i$	Linear operator that maps the fluorophore yield to measurements due to the i^{th} light source
$\tilde{\mathcal{B}}^i$	Finite element approximation of $\mathcal{B}^i$
$\mathcal{A}$	Linear forward operator that maps the fluorophore yield to all the measurements on the detectors
$\tilde{\mathcal{A}}$	Finite element approximation of $\mathcal{A}$
$\boldsymbol{x}$	Discretized fluorophore yield
$\boldsymbol{B}^i$	Discretized $\mathcal{B}^i$
$\boldsymbol{A}$	Discretized $\mathcal{A}$ - forward or sensing matrix
$\boldsymbol{\epsilon}$	Vector of measurement noise
σ_i	Variance of ϵ_i

continued...

continued from previous page

Symbol	Designation
$\boldsymbol{y}$	Vector of measurements contaminated by noise
λ	Regularization parameter
S	Sparsity of a signal
δ_S	S-restricted isometric constrant
$\mathcal{M}(\cdot)$	Mutual coherence of a matrix
$\mathcal{M}_1(\cdot,\cdot)$	Cumulative coherence of a matrix
$\boldsymbol{G}$	Matrix composed of the discretized Green's function of the emission light
$\boldsymbol{\Phi}$	Matrix composed of the discretized excitation light field
$r_{E_{p,q}}$	Normalized inner product between the p^{th} and q^{th} column of the matrix quantity E
$\boldsymbol{M}_A$	The preconditioning matrix of $\boldsymbol{A}$
$\boldsymbol{M}_d$	Digital masks on the detectors
$\boldsymbol{M}_s$	Optical masks on the point sources
Q_k	k^{th} structured illumination pattern
$\boldsymbol{G}_{new}$	Matrix composed of the discretized Green's function of the emission light field after the application of digital masks
$\boldsymbol{\Phi}_{new}$	Matrix composed of the discretized excitation light field after the application of the optical masks on sources
$\boldsymbol{A}_{pre}$	Forward/sensing matrix after the application of the preconditioning matrix $\boldsymbol{M}_A$
$\boldsymbol{A}_{new}$	Forward/sensing matrix after the application of the optical and digital masks
$\|\cdot\|_p$	l_p-norm of a vector quantity
$\|\cdot\|_F$	Frobenius norm of a matrix quantity

The rest of this chapter is organized as follows: Section 2.2 describes the light propagation model in FDOT; Section 2.3 and presents the FDOT forward and inverse problem formulation; Section 2.4 discretizes FDOT imaging problem by FEM; Section 2.5 presents the FDOT linear statistic model with additive noise; Finally, Section 2.6 concludes the chapter.

2.2 Light Propagation Model for FDOT

In this work, we use diffusion approximation of radiative transfer equation to model photon propagation in the medium. In the frequency domain, the excitation and emission light propagation in a bounded region $\Omega \in \mathbb{R}^3$ are modeled by two coupled diffusion equations [5]:

$$\begin{aligned} -\nabla \cdot D_x(\boldsymbol{r})\nabla\phi_x^i(\boldsymbol{r},\omega) + \left(\mu_{ax}(\boldsymbol{r}) + \frac{j\omega}{c}\right)\phi_x^i(\boldsymbol{r},\omega) &= s^i(\boldsymbol{r},\omega), \quad \boldsymbol{r} \in \Omega, \qquad (2.1) \\ -\nabla \cdot D_m(\boldsymbol{r})\nabla\phi_m^i(\boldsymbol{r},\omega) + \left(\mu_{am}(\boldsymbol{r}) + \frac{j\omega}{c}\right)\phi_m^i(r,\omega) &= \phi_x^i(\boldsymbol{r},\omega)\eta\mu_{axf}(\boldsymbol{r})\frac{1 - j\omega\tau(\boldsymbol{r})}{1 + (\omega\tau(\boldsymbol{r}))^2}, \\ &\quad \boldsymbol{r} \in \Omega, \qquad (2.2) \end{aligned}$$

where the subscripts x and m denote the excitation and emission wavelengths, respectively.

For the rest of our discussion, $E_{x,m}$ is a shorthand notation for the quantity E at either excitation or emission wavelength. $\boldsymbol{r} = [r_1, r_2, r_3] \in \Omega$ denotes the location in the medium; ω denotes the angular modulation frequency; $s^i(\boldsymbol{r},\omega)$ is the ith light source of the modulation frequency ω; $\phi_{x,m}^i(\boldsymbol{r},\omega)$ stands for the spatially varying photon density field at position $\boldsymbol{r}$ due to the source $s^i(\boldsymbol{r},\omega)$; c is the speed of light propagating in the medium; $D_{x,m}(\boldsymbol{r})$ stands for the diffusion coefficient in the medium at the excitation or emission wavelength; $\mu_{ax,m}(\boldsymbol{r})$ stands for the absorption coefficient of the medium at the excitation or emission wavelength; $\mu_{axf}(\boldsymbol{r})$ is the absorption coefficient of fluorophore at the excitation wavelength; η is the quantum yield and is defined as the ratio of the number of photons emitted to the number of photons absorbed; $\eta\mu_{axf}(\boldsymbol{r})$ is the fluorophore yield; $\tau(\boldsymbol{r})$ is the fluorescence lifetime.

The optical absorption coefficients μ_{ax} and μ_{am} are due to both the endogenous chromophores and exogenous fluorophore. Thus,

$$\mu_{ax}(\boldsymbol{r}) = \mu_{axe}(\boldsymbol{r}) + \mu_{axf}(\boldsymbol{r}), \qquad (2.3)$$

$$\mu_{am}(\boldsymbol{r}) = \mu_{ame}(\boldsymbol{r}) + \mu_{amf}(\boldsymbol{r}), \qquad (2.4)$$

where the subscript e denotes the endogenous chromophores, and the subscript f denotes the exogenous fluorophore.

We use the Robin-type boundary conditions given as follows [2]:

$$\begin{aligned} 2D_x(\boldsymbol{r})\hat{n}\cdot\nabla\phi_x^i(\boldsymbol{r},\omega)+\rho\phi_x^i(\boldsymbol{r},\omega) &= 0, \quad \boldsymbol{r}\in\partial\Omega, && (2.5)\\ 2D_m(\boldsymbol{r})\hat{n}\cdot\nabla\phi_m^i(\boldsymbol{r},\omega)+\rho\phi_m^i(\boldsymbol{r},\omega) &= 0, \quad \boldsymbol{r}\in\partial\Omega, && (2.6)\end{aligned}$$

where $\partial\Omega$ denotes the boundary of the region Ω, $\hat{n}$ denotes the outward normal of the boundary $\partial\Omega$, and ρ is the boundary mismatch parameter due to the light reflection at the boundary given as:

$$\rho=(1-R)/(1+R), \tag{2.7}$$

where R is a parameter governing the internal reflection at $\partial\Omega$. It can be approximated by :

$$R\approx -1.4399a^{-2}+0.7099a^{-1}+0.6681+0.0636a, \tag{2.8}$$

where a is the refractive index inside Ω.

In this thesis, our focus is to reconstruct the spatially varying fluorophore yield $\eta\mu_{axf}(\boldsymbol{r})$, given the boundary measurements of both the excitation and emission photon density fields. We made the following assumptions to simplify the problem:

1. The diffusion coefficients $D_{x,m}(\boldsymbol{r})$ at both the excitation and emission wavelengths are independent of the endogenous and exogenous absorption coefficients. Thus,

   $$D_{x,m}=\frac{1}{3(\mu'_{sx,m}(\boldsymbol{r})+\mu_{ax,m}(\boldsymbol{r}))}\approx\frac{1}{3\mu'_{sx,m}}, \tag{2.9}$$

   where $\mu_{sx,m}$ denotes the reduced scattering coefficient at the excitation or emission wavelength. Furthermore, $D_{x,m}(\boldsymbol{r})$ are known, and can be spatially varying over the imaging domain Ω.

2. The endogenous chromophores absorption coefficients $\mu_{axe}(\boldsymbol{r})$ and $\mu_{axe}(\boldsymbol{r})$ at both the excitation and emission wavelengths are known, and can be spatially varying over the imaging domain Ω.

3. The fluorescence life time $\tau(\boldsymbol{r})$ is known.

4. We use a single modulation frequency measurement model. A multi-frequency measurement model is a straightforward extension of the single-frequency measurement model. Furthermore, for notational brevity, we use $\omega = 0$ for the rest of our discussion. Thus (2.1-2.2) can be rewritten as follows:

$$\begin{aligned} -\nabla \cdot D_x(\boldsymbol{r})\nabla\phi_x^i(\boldsymbol{r}) + \mu_{ax}(\boldsymbol{r})\phi_x^i(\boldsymbol{r}) &= s^i(\boldsymbol{r}), \ \ \boldsymbol{r} \in \Omega, & (2.10)\\ -\nabla \cdot D_m(\boldsymbol{r})\nabla\phi_m^i(\boldsymbol{r}) + \mu_{am}(\boldsymbol{r})\phi_m^i(\boldsymbol{r}) &= \phi_x^i(\boldsymbol{r})\eta\mu_{axf}(\boldsymbol{r}), \ \ \boldsymbol{r} \in \Omega, & (2.11)\end{aligned}$$

and the Robin-type boundary condition in (2.5-2.6) can be rewritten as:

$$\begin{aligned} 2D_x(\boldsymbol{r})\hat{n} \cdot \nabla\phi_x^i(\boldsymbol{r}) + \rho\phi_x^i(\boldsymbol{r}) &= 0, \ \ \boldsymbol{r} \in \partial\Omega, & (2.12)\\ 2D_m(\boldsymbol{r})\hat{n} \cdot \nabla\phi_m^i(\boldsymbol{r}) + \rho\phi_m^i(\boldsymbol{r}) &= 0, \ \ \boldsymbol{r} \in \partial\Omega. & (2.13)\end{aligned}$$

In order to simplify the analysis later on, we make use of the adjoint problem associated with (2.11) and (2.13):

$$-\nabla \cdot D_m(\boldsymbol{r})\nabla g_m^j(\boldsymbol{r}) + \mu_{am}(\boldsymbol{r})g_m^j(\boldsymbol{r}) = 0, \quad \boldsymbol{r} \in \Omega \tag{2.14}$$

$$2D_m(\boldsymbol{r})\frac{\partial g_m^j(\boldsymbol{r})}{\partial n} + \rho g_m^j(\boldsymbol{r}) = s^{*j}(\boldsymbol{r}), \quad \boldsymbol{r} \in \partial\Omega \tag{2.15}$$

where $g_m^j(\boldsymbol{r})$ is the solution of the adjoint problem due to the jth adjoint point source s^{*j}. located at the jth detector position. We use the notation $g_m(\boldsymbol{r})$ to denote the Green's function of (2.11) and (2.13), and $g_m^j(\boldsymbol{r})$ is the shorthand notation of the Green's function of the emission light field evaluated at the j^{th} detector.

The absorption coefficients $\mu_{ax,m}$ of the medium is composed of the endogenous absorption $\mu_{aex,m}$ and the exogenous absorption $\mu_{afx,m}$ of the fluorophore:

$$\mu_{ax,m}(\boldsymbol{r}) = \mu_{aex,m}(\boldsymbol{r}) + \mu_{afx,m}(\boldsymbol{r}). \tag{2.16}$$

Since the excitation and emission wavelengths are typically close to each other, in many applications, the absorption coefficients at both wavelengths are assumed to be approximately equal [6]. Hence, we set

$$\mu_a \approx \mu_{ax} \approx \mu_{am}, \quad \mu_{af} \approx \mu_{axf} \approx \mu_{amf}. \tag{2.17}$$

2.3 FDOT Forward and Inverse Problems

2.3.1 Measurement Model

We assume that there are N_s light sources and N_d detectors located at the boundary of the imaging domain. From (2.10) and (2.12), the excitation light field due to the i^{th} source is given by

$$\phi_x^i(\boldsymbol{r}) = \int_\Omega g_x(\boldsymbol{r}, \boldsymbol{r'}) s^i(\boldsymbol{r'}) d\boldsymbol{r'}, \quad i = 1, ..., N_s \tag{2.18}$$

where g_x is the Green's function of (2.10) and (2.12). The excitation light interacts with the fluorophore inside the imaging domain and generates the emission light with a slightly longer wavelength

$$\phi_m^i(\boldsymbol{r}) = \int_\Omega g_m(\boldsymbol{r}, \boldsymbol{r'}) \phi_x^i(\boldsymbol{r'}) \eta \mu_{af}(\boldsymbol{r'}) d\boldsymbol{r'}, \quad i = 1, ..., N_s \tag{2.19}$$

Note that $\phi_x^i \eta \mu_{af}$ acts as the light source for the photon density field at the emission wavelength. Thus, we write the measurement $\Gamma_{i,j}$ at the j^{th} detector due to the i^{th} source as

$$\Gamma_{i,j} = \int_\Omega g_m^j(\boldsymbol{r}) \phi_x^i(\boldsymbol{r}) \eta \mu_{af}(\boldsymbol{r}) d\boldsymbol{r}, \quad i = 1, ..., N_s, \qquad i = 1, ..., N_d. \tag{2.20}$$

In general, (2.23) is nonlinear, since both g_m^j and ϕ_x^i depend on μ_a, which in turn, depends on μ_{af}.

2.3.2 Linearization of the FDOT Imaging Problem

2.3.2.1 Weak Fluorophore Linear Approximation

However, for the case of weak fluorophore [103], (2.23) can be linearized by assuming that the contribution of the fluorophore absorption to the total absorption is negligible, i.e.,

$$\mu_a = \mu_{ae} + \mu_{af} \approx \mu_{ae}. \tag{2.21}$$

Using (2.21), (2.23) can be linearized as

$$\Gamma_{i,j} \quad = \quad \int_\Omega g^j_{em}(\boldsymbol{r})\phi^i_{ex}(\boldsymbol{r})\eta\mu_{af}(\boldsymbol{r})d\boldsymbol{r} \tag{2.22}$$

where g^j_{em} is the Green's function of (2.2) and (2.6), and ϕ^i_{ex} is the excitation field when $\mu_a \approx \mu_{ae}$.

2.3.2.2 Nonlinear Iterative Perturbation

When strong fluorophore presents in the imaging domain, we iteratively linearize (2.23) to reconstruct $\eta\mu_{af}$. In each iteration, the green's function $g^{j^{(k)}}_m$ and the excitation light field $\phi^{i^{(k)}}_x$ are updated based on the estimated absorption coefficient $\mu^{(k-1)}_{af}$ of the pervious iteration, where the superscript k and $k-1$ denotes the number of iterations. (2.23) can be approximately reformulated as follows:

$$\delta\Gamma^{(k)}_{i,j} = \int_\Omega g^{j^{(k)}}_m(\boldsymbol{r})\phi^{i^{(k)}}_x(\boldsymbol{r})\eta\delta\mu^{(k)}_{af}(\boldsymbol{r})d\boldsymbol{r}, \quad i = 1, ..., N_s, \qquad i = 1, ..., N_d, \tag{2.23}$$

where $\delta\Gamma^{(k)}_{i,j}$ is the difference between the true and the estimated boundary measurement, $\eta\delta\mu^{(k)}_{af}$ is the perturbation between the estimated fluorophore yield from the previous iteration and the true fluorophore yield. From (2.23), $\eta\delta\mu^{(k)}_{ax,mf}$ is linearly mapped $\delta\Gamma^{(k)}_{i,j}$.

The estimated fluorophore yield at the kth iteration is given by

$$\mu^{(k)}_{ax,mf} = \delta\mu^{(k)}_{ax,mf} + \mu^{(k-1)}_{ax,mf}. \tag{2.24}$$

In the following discussion throughout this thesis, we assume weak fluorophore in the imaging domain, and use the linearization approach.

2.3.3 Forward Operator Formulation

We assume that there are N_s light sources and N_d detectors at the boundary of the imaging domain. Let $\mathbf{\Gamma}^i \in \mathbb{R}^{N_d}$ denote the vector of measurements due to the i^{th} source:

$$\mathbf{\Gamma}^i = [\Gamma_{i,1}, \Gamma_{i,2}, ..., \Gamma_{i,N_d}]^T.$$

We define the operator $\mathcal{B}^i$: $L(\Omega) \rightarrow \mathbb{R}^{N_d}$ as

$$(\mathcal{B}^i \mu)_j := \int_\Omega g^j_{em}(\boldsymbol{r}) \phi^i_{ex}(\boldsymbol{r}) \mu(\boldsymbol{r}) d\boldsymbol{r} = \int_\Omega b^i_j(\boldsymbol{r}) \mu(\boldsymbol{r}) d\boldsymbol{r} \tag{2.25}$$

where $\mu = \eta\mu_{af}$ denotes the fluorophore yield, and $b^i_j := g^{(j)}_{em}\phi^i_{ex}$ is the kernel of $\mathcal{B}^i$. Note that $\mathcal{B}^i$ linearly maps the fluorophore yield μ to the measurement vector $\mathbf{\Gamma}^i$:

$$\mathbf{\Gamma}^i = \mathcal{B}^i \mu. \tag{2.26}$$

We organize all the measurements $\mathbf{\Gamma}^i$, $i = 1, ..., N_s$ into a vector $\mathbf{\Gamma}$ of length $M = N_s \times N_d$

$$\mathbf{\Gamma} = [\mathbf{\Gamma}^{1^T}, ..., \mathbf{\Gamma}^{N_s{}^T}]^T \in \mathbb{R}^M. \tag{2.27}$$

Finally, we define the operator $\mathcal{A}$: $L(\Omega) \rightarrow \mathbb{R}^M$ which maps the fluorophore yield μ to $\mathbf{\Gamma}$ as

$$\begin{aligned} \mathbf{\Gamma} &= [\mathcal{B}^1, ..., \mathcal{B}^{N_s}]^T \mu \\ &=: \mathcal{A}\mu. \end{aligned} \tag{2.28}$$

2.4 Discretization of the FDOT Imaging Problem

2.4.1 Forward Problem Discretization

In this section, we calculate the excitation photon density field by FEM. Our calculation is based on the linearization approach, which can be easily extended to the nonlinear iterative perturbation approach. FEM offers the flexibility to deal with complex geometry and inhomogeneous distribution of absorption or diffusion coefficients [33]. In the forward problem discretization, we assume that the domain Ω is divided into P elements jointed at R vertex nodes. We seek for the finite element approximation of the $\phi_{ex}^{i}(\boldsymbol{r})$ and $g_{em}^{j}(\boldsymbol{r})$, which leads to the solution of the forward operator $\mathcal{B}$ and $\mathcal{A}$.

First, we solve for $\phi_{ex}^{i}(\boldsymbol{r})$. Using the first order Lagrange shape function, $\phi_{ex}^{i}(\boldsymbol{r})$ could be approximated by a continuous and piecewise linear function $\tilde{\phi}_{ex}^{i}(\boldsymbol{r})$ as follows:

$$\tilde{\phi}_{ex}^{i}(\boldsymbol{r}) = \sum_{k=1}^{R} p_k^i \psi_k(\boldsymbol{r}) = \boldsymbol{\psi}(\boldsymbol{r})^T \boldsymbol{p}^i, \tag{2.29}$$

where p_k^i is the kth nodal value of $\phi_{ex}^{i}(\boldsymbol{r})$, and $\boldsymbol{p}^i = [p_1^i, p_2^i, ..., p_R^i]^T$; $\psi_k(\boldsymbol{r})$ is the first order Lagrange basis function at node k, and $\boldsymbol{\psi}(\boldsymbol{r}) = [\psi_1(\boldsymbol{r}), \psi_2(\boldsymbol{r}), ..., \psi_R(\boldsymbol{r})]^T$.

The weak formulation of (2.1) is given by

$$\int_{\Omega} \chi(\boldsymbol{r})[-\nabla \cdot D_x(\boldsymbol{r})\nabla\tilde{\phi}_{ex}^{i}(\boldsymbol{r}) + \mu_{ax}(\boldsymbol{r})\tilde{\phi}_{ex}^{i}(\boldsymbol{r})d\Omega - s^i(\boldsymbol{r})]d^3\boldsymbol{r} = 0, \tag{2.30}$$

where $\chi(\boldsymbol{r})$ can be an arbitrary function that belongs to the Sobolev space $H^1(\Omega)$. The problem of solving for $\phi_{ex}^{i}(\boldsymbol{r})$ in (2.1) is now equivalent to finding a solution $\tilde{\phi}_{ex}^{i}(\boldsymbol{r})$ in (2.30), for all $\chi(\boldsymbol{r}) \in H^1(\Omega)$. In FEM, (2.30) is solvable when $\chi(\boldsymbol{r})$ can be represented by the shape functions $\psi_k, \;\; k = 1, 2, ..., R$.

$$\chi(\boldsymbol{r}) = \sum_{k=1}^{R} l_k \psi_k(\boldsymbol{r}) = \boldsymbol{l}^T \boldsymbol{\psi}(\boldsymbol{r}), \tag{2.31}$$

where l_k is the value of $\chi(\boldsymbol{r})$ at the kth node, and $\boldsymbol{l} = [l_1, l_2, ..., l_R]^T$. Now, the weak form in (2.30) can be written as follows:

$$\boldsymbol{l}^T \left[\int_{\Omega} \left(\boldsymbol{\psi}(\boldsymbol{r})(-\nabla \cdot D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}^T(\boldsymbol{r}))\boldsymbol{p}^i + \boldsymbol{\psi}(\boldsymbol{r})\mu_{ax}(\boldsymbol{r})\boldsymbol{\psi}^T(\boldsymbol{r})\boldsymbol{p}^i - \boldsymbol{\psi}(\boldsymbol{r})s^i(\boldsymbol{r}) \right) d^3\boldsymbol{r} \right] = 0. \tag{2.32}$$

Since $\boldsymbol{l}^T$ can be an arbitrary vector, (2.32) is equivalent to

$$\int_{\Omega} \boldsymbol{\psi}(\boldsymbol{r})(-\nabla \cdot D_x(\boldsymbol{r})\nabla\boldsymbol{p}^{i^T}(\boldsymbol{r})d^3\boldsymbol{r}\boldsymbol{p}^i + \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{r})\mu_{ax}(\boldsymbol{r})\boldsymbol{\psi}^T(\boldsymbol{r})d^3\boldsymbol{r}\boldsymbol{\Phi}_x = \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{r})s^i(\boldsymbol{r})d^3\boldsymbol{r}. \tag{2.33}$$

Using the following identity:

$$\nabla \cdot (f\;g) = f\nabla g + g\nabla f, \tag{2.34}$$

we get

$$\nabla \cdot \big(D_x(\boldsymbol{r})\boldsymbol{\psi}(\boldsymbol{r})\nabla\boldsymbol{\psi}^T(\boldsymbol{r})\big) = \boldsymbol{\psi}(\boldsymbol{r})\nabla \cdot \big(D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}^T(\boldsymbol{r})\big) + D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}(\boldsymbol{r}) \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r}), \tag{2.35}$$

and applying the divergence theorem

$$\int_\Omega \nabla \cdot \big(D_x(\boldsymbol{r})\boldsymbol{\psi}(\boldsymbol{r})\nabla\boldsymbol{\psi}^T(\boldsymbol{r})\big)\, d^3\boldsymbol{r} = \int_{\partial\Omega} D_x(\boldsymbol{r})\boldsymbol{\psi}(\boldsymbol{r})\hat{n} \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r})ds, \tag{2.36}$$

we can rewrite the first term in equation (2.33) as follows:

$$\begin{aligned}\int_\Omega \boldsymbol{\psi}(\boldsymbol{r})(-\nabla \cdot D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}^T(\boldsymbol{r}))d^3\boldsymbol{r}\boldsymbol{p}^i = &\int_{\partial\Omega} -D_x(\boldsymbol{r})\boldsymbol{\psi}(\boldsymbol{r})\hat{n} \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r})ds\boldsymbol{p}^i \\ &+ \int_\Omega D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}(\boldsymbol{r}) \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r})d^3\boldsymbol{r}\boldsymbol{p}^i.\end{aligned} \tag{2.37}$$

Recall from the boundary condition in (2.12),

$$D_x(\boldsymbol{r})\hat{n} \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r})\boldsymbol{p}^i = -\frac{\rho}{2}\boldsymbol{\psi}^T(\boldsymbol{r})\boldsymbol{p}^i, \quad \boldsymbol{r} \in \partial\Omega, \tag{2.38}$$

substituting (2.38) into (2.37), we get

$$\begin{aligned}\int_\Omega \boldsymbol{\psi}(\boldsymbol{r})(-\nabla \cdot D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}^T(\boldsymbol{r}))d^3\boldsymbol{r}\boldsymbol{p}^i = &\int_{\partial\Omega} \frac{\rho}{2}\boldsymbol{\psi}(\boldsymbol{r})\boldsymbol{\psi}^T(\boldsymbol{r})ds\boldsymbol{p}^i \\ &+ \int_\Omega D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}(\boldsymbol{r}) \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r})d^3\boldsymbol{r}\boldsymbol{p}^i.\end{aligned} \tag{2.39}$$

Substituting (2.39) back into (2.33), we get

$$\begin{aligned}\Bigg[\int_\Omega D_x(\boldsymbol{r})\nabla\boldsymbol{\psi}(\boldsymbol{r}) \cdot \nabla\boldsymbol{\psi}^T(\boldsymbol{r})d^3\boldsymbol{r} + \int_\Omega \boldsymbol{\psi}(\boldsymbol{r})\mu_{ax}(\boldsymbol{r})\boldsymbol{\psi}^T(\boldsymbol{r})d^3\boldsymbol{r} \\ + \int_{\partial\Omega} \frac{\rho}{2}\boldsymbol{\psi}(\boldsymbol{r})\boldsymbol{\psi}^T(\boldsymbol{r})ds\Bigg]\boldsymbol{p}^i = \int_\Omega \boldsymbol{\psi}(\boldsymbol{r})s^i(\boldsymbol{r})d^3\boldsymbol{r}.\end{aligned} \tag{2.40}$$

The (2.40) can be rewritten in the matrix form as follows:

$$(\boldsymbol{K} + \boldsymbol{C} + \frac{\rho}{2}\boldsymbol{F})\boldsymbol{p}^i = \boldsymbol{u}^i \tag{2.41}$$

where $\boldsymbol{K}$ is an $R \times R$ matrix,

$$\boldsymbol{K} = \int_\Omega D_x(\boldsymbol{r}) \nabla \boldsymbol{\psi}(\boldsymbol{r}) \cdot \nabla \boldsymbol{\psi}^T(\boldsymbol{r}) d^3\boldsymbol{r}, \tag{2.42}$$

with the (t,k)th entry given by

$$K_{t,k} = \int_\Omega D_x(\boldsymbol{r}) \nabla \psi_t(\boldsymbol{r}) \nabla \psi_k(\boldsymbol{r}) d^3\boldsymbol{r}. \tag{2.43}$$

$\boldsymbol{C}$ is an $R \times R$ matrix,

$$\boldsymbol{C} = \int_\Omega \boldsymbol{\psi}(\boldsymbol{r}) \mu_{ax}(\boldsymbol{r}) \boldsymbol{\psi}^T(\boldsymbol{r}) d^3\boldsymbol{r}, \tag{2.44}$$

where the (t,k)th entry is given by

$$C_{t,k} = \int_\Omega \mu_{ax}(\boldsymbol{r}) \psi_t(\boldsymbol{r}) \psi_k(\boldsymbol{r}) d^3\boldsymbol{r}. \tag{2.45}$$

$\boldsymbol{F}$ is an $R \times R$ matrix,

$$\boldsymbol{F} = \int_{\partial\Omega} \frac{\rho}{2} \boldsymbol{\psi}(\boldsymbol{r}) \boldsymbol{\psi}^T(\boldsymbol{r}) ds, \tag{2.46}$$

where the (t,k)th entry is given by

$$F_{t,k} = \int_{\partial\Omega} \psi_t(\boldsymbol{r}) \psi_k(\boldsymbol{r}) ds. \tag{2.47}$$

$\boldsymbol{u}^i$ is a vector of length R,

$$\boldsymbol{u}^i = \int_\Omega \boldsymbol{\psi}(\boldsymbol{r}) s^i(\boldsymbol{r}) d^3\boldsymbol{r}, \tag{2.48}$$

where the tth entry is given by

$$u_t^i = \int_\Omega s^i(\boldsymbol{r}) \psi_t(\boldsymbol{r}) d^3\boldsymbol{r}. \tag{2.49}$$

Then, $\boldsymbol{p}^i$ in (2.41) can be solved by matrix inversion:

$$\boldsymbol{p}^i = (\boldsymbol{K} + \boldsymbol{C} + \frac{\rho}{2}\boldsymbol{F})^{-1} \boldsymbol{u}^i, \tag{2.50}$$

Combining (2.50) and (2.29), we obtain the excitation photon density field as follows:

$$\tilde{\phi}^i_{ex}(\boldsymbol{r}) = \boldsymbol{\psi}^T(\boldsymbol{r})\boldsymbol{p}^i = \boldsymbol{\psi}^T(\boldsymbol{r})(\boldsymbol{K} + \boldsymbol{C} + \frac{\rho}{2}\boldsymbol{F})^{-1}\boldsymbol{u}^i \tag{2.51}$$

The Green's function $g^j_{em}(\boldsymbol{r})$ can be calculated in a similar way as $\phi^i_{ex}(\boldsymbol{r})$. We first approximate $g^j_{em}(\boldsymbol{r})$ by a continuous and piecewise linear function $\tilde{g}^j_{em}(\boldsymbol{r})$ as follows:

$$\tilde{g}^j_{em}(\boldsymbol{r}) = \sum_{k=1}^{R} q_k\psi_k(\boldsymbol{r}) = \boldsymbol{\psi}^T(\boldsymbol{r})\boldsymbol{q}, \tag{2.52}$$

where q_k is the kth nodal value of $g^j_{em}(\boldsymbol{r})$, and $\boldsymbol{q} = [q_1, ..., q_R]^T$. For convenience, we use the same element shape functions $\psi_k,\ k = 1, 2, ..., R$ as that if the excitation field in (2.29). However, note that the finite element shape functions do not need to be the same for both the excitation and emission fields. Similar as the excitation light field, we use the adjoint problem formulation in (2.14) and (2.15), and solve for $\boldsymbol{q}$.

The finite element approximation of $b^i_j(\boldsymbol{r})$ in (2.25) is given by

$$\tilde{b}^i_j = \tilde{g}^j_{em}(\boldsymbol{r})\tilde{\phi}^i_{ex}(\boldsymbol{r}). \tag{2.53}$$

Thus, the finite element approximation of the forward operator $\mathcal{B}^i$ and $\mathcal{A}$ are given by

$$(\tilde{\mathcal{B}}^i\mu)_j := \int_\Omega \tilde{b}^i_j(\boldsymbol{r})\mu(\boldsymbol{r})d\boldsymbol{r}, \tag{2.54}$$

and

$$\tilde{\mathcal{A}} = [\tilde{\mathcal{B}}^1, \tilde{\mathcal{B}}^2, ..., \tilde{\mathcal{B}}^{N_s}]^T. \tag{2.55}$$

2.4.2 Inverse Problem Discretization

In this work, our goal is to reconstruct the fluorophore yield $\mu(\boldsymbol{r})$ in the imaging domain Ω. We decompose the imaging domain Ω into Q elements jointed at N vertex nodes. The fluorophore yield $\mu(\boldsymbol{r})$ can be approximated by its finite element counterparts $\tilde{\mu}(\boldsymbol{r})$:

$$\tilde{\mu}(\boldsymbol{r}) = \sum_{k=1}^{N} x_k\varphi_k(\boldsymbol{r}) = \boldsymbol{x}^T\boldsymbol{\varphi}(\boldsymbol{r}) \tag{2.56}$$

where $\varphi_k(\boldsymbol{r})$ is the kth Lagrange shape function, $\boldsymbol{\varphi}(\boldsymbol{r}) = [\varphi_1(\boldsymbol{r}), ..., \varphi_N(\boldsymbol{r})]^T$, and x_k is the fluorophore yield coefficient at the kth vertex node, $\boldsymbol{x} = [x_1, ..., x_N]^T$. Similarly, $\tilde{\tilde{b}}_j^i$ is the finite element approximation of $\tilde{b}_j^i$,

$$\tilde{\tilde{b}}_j^i(\boldsymbol{r}) = \sum_{k=1}^{N} v_k \varphi_k(\boldsymbol{r}) = \boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{v} \tag{2.57}$$

where v_k is the kth vertex node of $\tilde{\tilde{b}}_j^i$, given by

$$v_k = g_{em,k}^j \phi_{ex,k}^i. \tag{2.58}$$

$g_{em,k}^j$ is the discretized Green's function evaluated at the j^{th} detector due to the k^{th} voxel; and $\phi_{ex,k}^i$ is the discretized excitation light field at the k^{th} voxel due to the i^{th} source. Substitute (2.56) and (2.57) back into (2.25), we have

$$\begin{aligned} \Gamma_{i,j} &= \int_\Omega \tilde{\tilde{b}}_j^i(\boldsymbol{r}) \mu(\boldsymbol{r}) d\boldsymbol{r} \\ &= \int_\Omega \boldsymbol{v}^T \boldsymbol{\varphi}(\boldsymbol{r}) \boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{x} \\ &= \boldsymbol{v}^T \boldsymbol{x}. \end{aligned} \tag{2.59}$$

The last equality comes from the orthonormal property of the Lagrange shape function,

$$\int_\Omega \varphi_{k_1}(\boldsymbol{r}) \varphi_{k_2}(\boldsymbol{r}) d\boldsymbol{r} = \begin{cases} 1 & k_1 = k_2 \\ 0 & k_1 \neq k_2 \end{cases}. \tag{2.60}$$

Thus, from (2.26) (2.58) and (2.59), we obtain:

$$\boldsymbol{\Gamma}^i = \boldsymbol{B}^i \boldsymbol{x}, \tag{2.61}$$

where

$$\boldsymbol{B}^i = \begin{bmatrix} g_{em,1}^1 \phi_{ex,1}^i & \cdots & g_{em,N}^1 \phi_{ex,N}^i \\ \vdots & & \vdots \\ g_{em,1}^{N_d} \phi_{ex,1}^i & \cdots & g_{em,N}^{N_d} \phi_{ex,N}^i \end{bmatrix} \in \mathbb{R}^{N_d \times N} \tag{2.62}$$

Taking into account all the measurements due to all the light sources, the discretized form of (2.28) becomes

$$\mathbf{\Gamma} = \boldsymbol{A}\boldsymbol{x}, \tag{2.63}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B}^1 \\ \boldsymbol{B}^2 \\ \vdots \\ \boldsymbol{B}^{N_s} \end{bmatrix} = \begin{bmatrix} g^1_{em,1}\phi^1_{ex,1} & \cdots & g^1_{em,N}\phi^1_{ex,N} \\ \vdots & & \vdots \\ g^{N_d}_{em,1}\phi^1_{ex,1} & \cdots & g^{N_d}_{em,N}\phi^1_{ex,N} \\ g^1_{em,1}\phi^2_{ex,1} & \cdots & g^1_{em,N}\phi^2_{ex,N} \\ \vdots & & \vdots \\ g^{N_d}_{em,1}\phi^{N_s}_{ex,1} & \cdots & g^{N_d}_{em,N}\phi^{N_s}_{ex,N} \end{bmatrix} \in \mathbb{R}^{M\times N} \tag{2.64}$$

is the vector-valued forward operator. Note that we refer to $\boldsymbol{A}$ as the *sensing or forward operator* interchangeably for the rest of the thesis.

2.5 Linear FDOT Statistical Model with Additive Noise

Now, the discretized version of the linear model in (2.28) is given as

$$\mathbf{\Gamma} = \boldsymbol{A}\boldsymbol{x}. \tag{2.65}$$

Consider the noisy measurements:

$$\boldsymbol{y} = \mathbf{\Gamma} + \boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{2.66}$$

where $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement data with the additive noise $\boldsymbol{\epsilon}$,

$$\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, ..., \epsilon_M]^T. \tag{2.67}$$

We model ϵ_i, $i = 1, 2, ..., M$, as independent Gaussian random variables, i.e.,

$$\epsilon_i \sim \mathcal{N}(0, \sigma_i^2), \tag{2.68}$$

where σ_i is the variance of ϵ_i.

2.6 Conclusion

In this chapter, we modeled the light propagation in the imaging domain by two coupled diffusion equations. We addressed the nonlinearity of the FDOT imaging problem by weak fluorophore assumption, which can be easily extend to an iterative perturbation approach when strong fluorophore is present. Next, we gave a brief description of the FDOT forward and inverse problem, and use FEM for discretization. Finally, we presented the formulation of the vector-valued forward operator, which can be used in the fluorophore map reconstruction in Chapter 3-6.

CHAPTER 3
REGULARIZATION METHODS IN FDOT

3.1 Introduction

In this chapter, we exploit various regularization techniques to reconstruct the sparse fluorophore map in the imaging domain. The focus of the FDOT inverse problem is to estimate the fluorophore yield in the imaging domain given the boundary measurements. One of the challenges in the FDOT inverse problem is that the forward sensing matrix $\boldsymbol{A}$ is highly ill-posed, in the sense that the number of fluorophore yield coefficients that we estimate is much larger than the number of measurements available. Thus, it is necessary to use regularization techniques to address the ill-posedness of the linear inverse problem.

A number of regularization techniques have been reported for the inverse problem of DOT or FDOT [37–44]. In this chapter, we give a brief introduction various state-of-art regularization techniques, and apply the regularization techniques to the FDOT inverse problem. The performances of different regularization techniques are quantitatively evaluated using synthetic data generated by 3D numerical simulation and real silicon phantom experiment.

In addition, we propose a combined l_2-l_p-norm regularization method for the FDOT inverse problem. In the FDOT inverse problem, the true fluorophore yield map is often piecewise constant, where the fluorophore is concentrated within a small foreground region. Thus, the fluorophore yield coefficients have large amplitudes in the foreground region, and are close to zero in the large background region. We note that the Tikhonov regularization method, which penalizes the l_2-norm of the fluorophore yield coefficients that we want to reconstruct, often leads to an over-smoothed solution [104]. One way to fix the problem of over-smoothing is to replace the quadratic penalty function with a function that increases less rapidly for large argument values, such as the l_p-norm penalty function where $1 \leq p < 2$. Unfortunately, l_p $(1 \leq p < 2)$ norm penalty function preserves large spikes of noise in the reconstructed image. A potential solution to this problem is to use a combined l_2-l_p-norm regularization. In this work, we examine the improvements in the image resolution and contrast as a result of application of the combined l_2-l_p-norm regularization in image reconstruction.

The rest of this chapter is organized as follows: Section 3.2 reviews a number of widely used regularization techniques; Section 3.3 presents the novel combined l_2-l_p-norm regularization method; Section 3.4 presents the numerical simulation; Section 3.5 shows the real phantom experiment results; Finally, Section 3.6 reaches the conclusion.

3.2 Linear Regularization Techniques in Literature

Recall from Chapter 2, we present a linear relationship between the measurements $\boldsymbol{y}$ and the fluorophore coefficients,

$$\boldsymbol{y} = \boldsymbol{\Gamma} + \boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}. \tag{3.1}$$

The linear system in (3.1) is ill-posed, thus regularization techniques are required to recovery $\boldsymbol{x}$ from $\boldsymbol{y}$.

3.2.1 Truncated Singular Value Decomposition

In linear algebra, the Singular Value Decomposition (SVD) is the generalization of the spectral factorization of an arbitrary matrix.

The forward matrix $\boldsymbol{A}$ is an $M \times N$ matrix, and the SVD of $\boldsymbol{A}$ is given by,

$$\boldsymbol{A} = \boldsymbol{U}_A \Sigma_A \boldsymbol{V}_A^T, \tag{3.2}$$

where $\boldsymbol{U}_A$ is an $M \times M$ unitary matrix, $\boldsymbol{V}_A$ is an $N \times N$ unitary matrix, the superscript T denotes the conjugate transpose, the matrix $\boldsymbol{\Sigma}_A$ is an $M \times N$ diagonal matrix whose diagonal elements are the singular values of $\boldsymbol{A}$,

$$\varrho_1 \geq \varrho_2 \geq \cdots \varrho_M \geq 0. \tag{3.3}$$

Let p, $1 \leq p \leq M$, be the largest index for which $\varrho_p > 0$. The minimum norm solution of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ is given as:

$$\boldsymbol{x} = \boldsymbol{V}_A \boldsymbol{\Sigma}_{\mathbf{A}}{}^{\dagger} \boldsymbol{U}_A^T \boldsymbol{y}, \tag{3.4}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{A}}{}^{\dagger} = \begin{bmatrix} 1/\varrho_1 & 0 & & \dots & & 0 \\ 0 & 1/\varrho_2 & & & & \\ & & \dots & & & \\ \vdots & & & 1/\varrho_p & & \vdots \\ & & & & 0 & \\ & & & & & \\ & & & & & \\ 0 & & & \dots & & 0 \end{bmatrix} \in \mathbb{R}^{N \times M}. \tag{3.5}$$

In practice, we have $\boldsymbol{y} \approx \boldsymbol{A}\boldsymbol{x}$, and the forward matrix $\boldsymbol{A}$ is ill-posed, with its smallest positive singular values very close to zero. Thus the minimum norm solution is very sensitive to measurement errors in the vector $\boldsymbol{y}$, and we often have to choose a truncation parameter $k \leq p$. We refer to the discrepancy principle to choose k, which states that the approximate solution can not yield a smaller residual error than the measurement noise level, since otherwise the solution would fit to the noise. The measurement vector $\boldsymbol{y}$ is a noisy approximation of the noiseless measurement vector $\mathbf{\Gamma}$, and the measurement noise level is given by:

$$\|\boldsymbol{y} - \mathbf{\Gamma}\|_2 = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 = \|\boldsymbol{\epsilon}\|_2. \tag{3.6}$$

From the discrepancy principle, we choose the truncation parameter k to be the largest index that satisfies:

$$\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_k\|_2^2 \leq \|\boldsymbol{\epsilon}\|_2^2 = \sum_{i=1}^{N} \sigma_i^2, \tag{3.7}$$

where $\boldsymbol{x}_k$ is the solution when the truncation parameter is k.

3.2.2 Conjugate Gradient Method

CG is an iterative solver of large linear equation systems [38, 105]. CG works well with sparse systems that are too large to be handled by matrix inversion methods such as Cholesky decomposition. It is guaranteed to find a solution if the forward matrix $\boldsymbol{A}$ is symmetric and positive definite. Thus, we first multiply both sides of (3.1) by $\boldsymbol{A}^T$:

$$\boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{\epsilon}, \tag{3.8}$$

using notation $\boldsymbol{y}'$ to denote $\boldsymbol{A}^T\boldsymbol{y}$, $\boldsymbol{A}'$ to denote $\boldsymbol{A}^T\boldsymbol{A}$, and $\boldsymbol{\epsilon}'$ to denote $\boldsymbol{A}^T\boldsymbol{\epsilon}$, we rewrite (3.8) as follows:

$$\boldsymbol{y}' = \boldsymbol{A}'\boldsymbol{x} + \boldsymbol{\epsilon}', \tag{3.9}$$

where $\boldsymbol{A}'$ is now a symmetric and positive definite $N \times N$ matrix.

The linearly independent vectors $\boldsymbol{s}_1, ..., \boldsymbol{s}_k$ are said to be $\boldsymbol{A}'$-conjugate, if

$$< \boldsymbol{s}_i, \boldsymbol{s}_j >_{\boldsymbol{A}} = \boldsymbol{s}_i^T\boldsymbol{A}'\boldsymbol{s}_j = 0, \text{ for } i \neq j. \tag{3.10}$$

In other words, the vectors $\boldsymbol{s}_i$, $i = 1, 2, ..., k$, are orthogonal to each other with respect to the inner product. Roughly, the idea of CG is to produce a sequence of approximate solutions as linear combinations of the vectors $\boldsymbol{s}_1, ..., \boldsymbol{s}_k$.

If we have an approximate solution $\boldsymbol{x}_k$ which is a linear combination of $\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{k-1}$,

$$\boldsymbol{x}_k = \sum_{i=1}^{k-1} \alpha_i \boldsymbol{s}_i, \tag{3.11}$$

where α_i is the coefficient of $\boldsymbol{s}_i$. The residual after the kth iteration is given by

$$\boldsymbol{e}_k = \boldsymbol{y}' - \boldsymbol{A}'\boldsymbol{x}_k, \tag{3.12}$$

where

$$\boldsymbol{e}_k \perp \text{span}\{\boldsymbol{s}_1, ..., \boldsymbol{s}_{k-1}\}. \tag{3.13}$$

We choose the next search direction $\boldsymbol{s}_k$ to be $\boldsymbol{A}'$-conjugate with the previous ones,

$$\boldsymbol{s}_k = \boldsymbol{e}_k + \beta_{k-1}\boldsymbol{s}_{k-1}, \tag{3.14}$$

where $\beta_{k-1} = \frac{\|\boldsymbol{r}_k\|^2}{\|\boldsymbol{r}_{k-1}\|^2}$. Hence, given a sequence $\{\boldsymbol{s}_k\}$ of directions, we produce a sequence $\boldsymbol{x}_k$ of the approximate solutions by setting

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{s}_k, \quad \alpha_k = \frac{\boldsymbol{s}_k^T \boldsymbol{e}_k}{\boldsymbol{s}_k^T \boldsymbol{A}' \boldsymbol{s}_k}, \quad \boldsymbol{e}_k = \boldsymbol{e}_{k-1} - \alpha_{k-1} \boldsymbol{A}' \boldsymbol{s}_{k-1}, \tag{3.15}$$

repeat this process until convergence.

From the discrepancy principle, the iterations are terminated as soon as the residual error is equal to or smaller than the measurement noise level.

The Pseudocode for Conjugate Gradient method is given in Algorithm 1:

Algorithm 1 Conjugate Gradient Method

```
input: $\boldsymbol{A}'$, $\boldsymbol{y}'$;
output: $\hat{\boldsymbol{x}}$
Pick $\boldsymbol{x}_1$. Set $k = 1$, $\boldsymbol{e}_1 = \boldsymbol{y}' - \boldsymbol{A}'\boldsymbol{x}_1$,$\boldsymbol{s}_1 = \boldsymbol{e}_1$;
Repeat until convergence:
    $\alpha_k = \| \boldsymbol{e}_k \|_2^2 / \boldsymbol{s}_k^T \boldsymbol{A}' \boldsymbol{s}_k$;
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{s}_k$;
    $\boldsymbol{e}_{k+1} = \boldsymbol{e}_k - \alpha_k \boldsymbol{A}' \boldsymbol{s}_k$;
    $\beta_k = \| \boldsymbol{e}_{k+1} \|_2^2 / \| \boldsymbol{e}_k \|_2^2$;
    $\boldsymbol{s}_{k+1} = \boldsymbol{e}_{k+1} + \beta_k \boldsymbol{s}_k$;
    $k \leftarrow k + 1$
end
$\hat{\boldsymbol{x}} = \boldsymbol{x}_k$;
```

3.2.3 Kaczmarz Iteration Method

The Kaczmarz iteration method is also known as the *algebraic reconstruction technique* (ART). The idea in the Kaczmarz iteration method is to successively project an estimated solution onto subspaces defined by the rows of the forward matrix $\boldsymbol{A}$.

Let $\boldsymbol{A}$ be partitioned as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \\ \boldsymbol{A}_l \end{bmatrix} \in \mathbb{R}^{M \times N}, \ \ \boldsymbol{A}_j \in \mathbb{R}^{k_j \times N}, \tag{3.16}$$

where $\boldsymbol{A}_j$ is of rank k_j, and $k_1 + k_2 + ... + k_l = M$. We consider the systems:

$$\boldsymbol{A}_j \boldsymbol{x} = \boldsymbol{y}_j, \quad 1 \leq j \leq l, \tag{3.17}$$

where $\boldsymbol{y}_j \in \mathbb{R}^{k_j}$. Let X_j denote the solution space of (3.17):

$$X_j = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}_j \boldsymbol{x} = \boldsymbol{y}_j\}, \tag{3.18}$$

and let $P_j : \mathbb{R}^N \rightarrow X_j$ denote the orthogonal projection of an arbitrary vector $\boldsymbol{s} \in \mathbb{R}^N$ onto X_j,

$$P_j \boldsymbol{s} = \boldsymbol{s} + \boldsymbol{A}_j^T (\boldsymbol{A}_j \boldsymbol{A}_j^T)^{-1} (\boldsymbol{y}_j - \boldsymbol{A}_j \boldsymbol{s}). \tag{3.19}$$

We further define the sequential projection:

$$P = P_l P_{l-1} \dots P_2 P_1. \tag{3.20}$$

Let $\boldsymbol{x}_k$ be the solution in the kth iteration, the Kaczmarz sequence $\{\boldsymbol{x}_k\}$ is defined recursively as

$$\boldsymbol{x}_{k+1} = P \boldsymbol{x}_k, \quad \boldsymbol{x}_0 = 0. \tag{3.21}$$

where $\boldsymbol{x}_k$ converges to the minimum norm solution of the equation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

In practice, we have $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon} \approx \boldsymbol{A}\boldsymbol{x}$. From the discrepancy principle, the iterations are terminated when the residual error $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2$ reaches the measurement noise level $\|\boldsymbol{\epsilon}\|_2$.

The Pseudocode of Kaczmarz iteration is given in Algorithm 2:

3.2.4 Tikhonov Regularization Method

When solving for the fluorophore yield coefficient vector $\boldsymbol{x}$, problems occur when the singular values of the matrix $\boldsymbol{A}$ are very close to zero, causing the norm of the approximate solution to go to infinity. The idea in the Tikhonov regularization scheme is to control simultaneously the square residue error $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ and the l_2-norm of the approximate solution $\|\boldsymbol{x}\|_2^2$, which seeks a minimizer of the function,

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_2^2, \tag{3.22}$$

Algorithm 2 Kaczmarz Iteration

```
input: $\boldsymbol{A}'$, $\boldsymbol{y}'$;
output: $\hat{\boldsymbol{x}}$
Set $k = 0$, $\boldsymbol{x}_0 = 0$;
Repeat until convergence:
    $\boldsymbol{z}_0 = \boldsymbol{x}_k$;
    for $j = 1$: $M$ repeat
        $\boldsymbol{z}_j = \boldsymbol{z}_{j-1} + (1/ \parallel \boldsymbol{A}_j \parallel_2^2)(\boldsymbol{y}_j - \boldsymbol{A}_j^T \boldsymbol{z}_{j-1})\boldsymbol{A}_j$;
    end
    $\boldsymbol{x}_{k+1} = \boldsymbol{z}_M$;
    $k \leftarrow k + 1$;
end
$\hat{\boldsymbol{x}} = \boldsymbol{x}_k$;
```

where the parameter $\lambda > 0$ is called the regularization parameter. A unique analytical solution of (3.22) is given by:

$$\boldsymbol{x}_\lambda \;=\; \arg\min_{\boldsymbol{x}} J(\boldsymbol{x}) \;=\; (\boldsymbol{A}^T\boldsymbol{A} + \lambda\mathbf{I})^{-1}\boldsymbol{A}^T\boldsymbol{y}. \tag{3.23}$$

We define the l_2-norm of the residual error as follows:

$$f(\lambda) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_\lambda\|_2^2, \tag{3.24}$$

where $f(\lambda)$ is a strictly increasing function in λ. The discrepancy principle states that the residual error should not be smaller than the measurement noise level, thus we choose the regularization parameter δ when the residual error equals to the noise level,

$$f(\lambda) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_\lambda\|_2^2 = \|\boldsymbol{\epsilon}\|_2^2. \tag{3.25}$$

3.2.5 The L_p-Norm $(1 \leq p < 2)$ Regularization

In practice, the Tikhonov regularization often tends to produce an over-smoothed solution, which is incapable of recovering discontinuities in the presence of noise. One scheme to address the problem of over-smoothing is to replace the quadratic penalty function in (3.22) with a non-quadratic function that increases less rapidly for sufficiently large arguments, such as the l_p-norm $(1 \leq p < 2)$ penalty function $\lambda_p\|\boldsymbol{x}\|_p$, $(1 \leq p < 2)$.

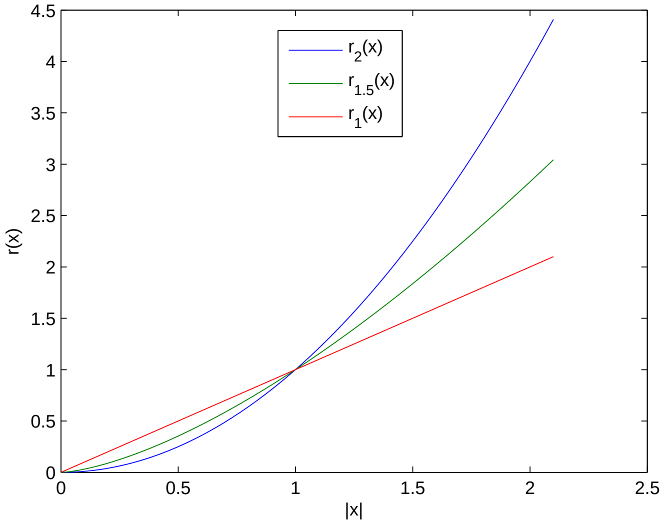


Figure 3.1: A comparison of the cost functions of the Tikhonov regularization and l_p-norm regularization.

Let

$$R_p(\boldsymbol{x}) = \lambda_{l_p}\|\boldsymbol{x}\|_p^p = \sum_{i=1}^{N} r_p(x_i) = \sum_{i=1}^{N} \lambda_{l_p}|x_i|^p \tag{3.26}$$

denotes the l_p $(1 \leq p < 2)$ norm penalty function with the regularization parameter λ_{l_p}, and $r_p(x_i) = \lambda_{l_p}|x_i|^p$ is the cost function for the ith fluorophore yield coefficient. Let

$$R_2(\boldsymbol{x}) = \lambda_{l_2}\|\boldsymbol{x}\|_2^2 = \sum_{i=1}^{N} r_2(x_i) = \sum_{i=1}^{N} \lambda_{l_2}|x_i|^2 \tag{3.27}$$

denotes the Tikhonov penalty function with the regularization parameter λ_{l_2}, and $r_2(x_i) = \lambda_{l_2}|x_i|^2$ is the cost function for the ith fluorophore yield coefficient.

When $|x_i| < (\frac{\lambda_{l_p}}{\lambda_{l_2}})^{p-2}$, we have $r_p(x_i) > r_2(x_i)$ $(1 \leq p < 2)$, thus for small arguments in the vector $\boldsymbol{x}$, l_p-norm regularization imposes more penalty than the Tikhonov regularization. On the other hand, when $|x_i| \geq (\frac{\lambda_{l_p}}{\lambda_{l_2}})^{p-2}$, we have $r_p(x_i) \leq r_2(x_i)$ $(1 \leq p < 2)$. Thus, for large arguments in the vector $\boldsymbol{x}$, l_p-norm regularization imposes less penalty than the the Tikhonov regularization. In addition, when $p \to 1$, the l_p-norm regularization leads to a sparse solution [106]. The solution can be obtained from Iterative Ridge Regression [107]. A comparison of the Tikhonov and the l_p-norm penalty functions is shown in Figure 3.1, where $\lambda_{l_2} = \lambda_{l_{1.5}} = \lambda_{l_1} = 1$. It is shown that for $|x| < 1$, $r_2(x) < r_{1.5}(x) < r_1(x)$, and for $|x| \geq 1$, $r_1(x) \leq r_{1.5}(x) \leq r_2(x)$.

In the FDOT inverse problem, the fluorophore is often concentrated within a small foreground region, where the fluorophore yield coefficients have large amplitudes, while in the large background region, the fluorophore yield coefficients are close to zero. In this case, the fluorophore yield coefficients that we want to reconstruct are sparse in the imaging domain. Therefore, the l_p-norm regularization is a natural choice for the FDOT inverse problem. Compared to the Tikhonov regularization, the l_p-norm regularization suppresses small values in the fluorophore yield coefficients due to the background noise, and preserves large values due to the fluorophore concentration.

3.3 The Combined L_2-L_p-Norm Regularization

3.3.1 *A Priori* Information in FDOT

In FDOT, the use of *a priori* information improves the FDOT reconstruction both visually and quantitatively [20, 108–116]. The most commonly used *a priori* information is the anatomical structure information obtained from another imaging modality, such as X-ray CT [108, 109], MRI [20, 110, 111] and others [112, 113]. These multi-modality developments are motivated by the fact that the contrast from high resolution anatomical imaging modalities correlate well with the optical properties [20, 108].

There are a number of approaches to incorporate structural *a priori* information to FDOT image reconstruction. The most straightforward one is to assume that the perturbations in optical properties are present only in small *regions of interest* (ROI) which can be determined by an anatomical image [114, 115]. A second approach involves segmenting the whole imaging domain into different types of regions according to anatomical *a priori* information and assuming that each region has homogeneous optical properties [116]. Both methods rely on accurate co-registration of the optical and anatomical images, and thus, are referred to as "*hard prior*".

In practice, it is possible that the anatomical *a priori* information does not correlate well with the optical properties. As a result, the "*hard prior*" approach can lead to erroneous and strongly biased images. As an alternative, "*soft prior*" methods are proposed to represent optical-anatomy correlation, where the anatomical *a priori* information usually interpreted as coefficients in the regularization term [117–119]. However, although "soft prior" approach was proposed to represent optical-anatomy correlation, where the anatomical *a priori* information usually included as parameters in the regularization term [117–119]. Compared with "hard prior" approach, "soft prior" approach introduces some flexibilities in dealing with the correlation between anatomical prior and optical properties.

In this section, we take an alternative approach in incorporating the prior information. Instead of characterizing the foreground region of fluorophore in the imaging domain, we use a *quantitative range* of the amplitudes of fluorophore coefficients in the foreground region is used as *a priori* information. We design the regularization term to incorporate the estimated amplitudes of fluorophore coefficients. The range of fluorophore coefficients can be simply obtained by first reconstructing the fluorophore yield using the Tikhonov regularization. This specific *a prior* information is able to quantitatively improve the FDOT reconstruction without any bias towards anatomical structure information. The detailed methods are described in the following subsection.

3.3.2 The Combined L_2-l_p-Norm Regularization with Specific *A Priori* Information

The l_p-norm regularization tends to preserve the fluorophore yield coefficients of large amplitudes, however, it also produces large spikes due to noise or the ill-conditioning of the forward matrix. Suppose we know *a prior* that the amplitudes of the real fluorophore coefficients in the foreground region are in the range $[\alpha,\ \beta]$. Thus if the reconstructed fluorophore yield coefficient at the ith voxel is too large, i.e., $x_i > x\beta$, it is probably due to the noise in the measurement or the ill-conditioning of the forward matrix $\boldsymbol{A}$. The drawback of the l_p-norm regularization is that it is less effective in smoothing out such large spikes compared to the Tikhonov regularization.

We propose a penalty function combining the l_p and l_2-norm regularization,

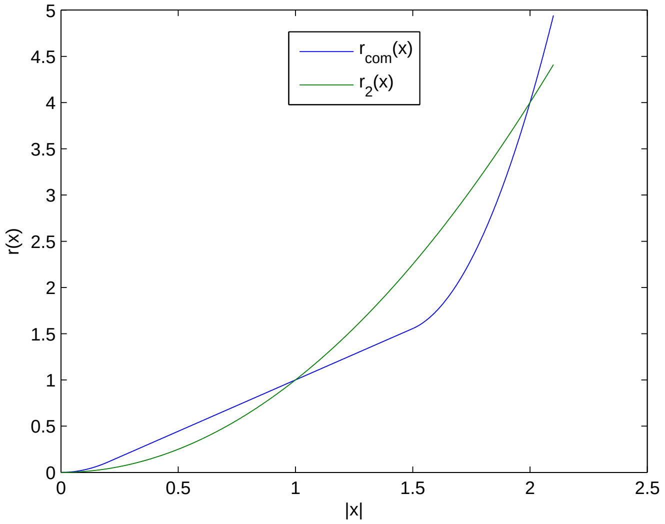


Figure 3.2: A comparison of the cost functions of the Tikhonov regularization and the combined l_2-l_p-norm ($p = 1$) regularization.

which imposes small penalty when x_i is within the range $[\alpha,\ \beta]$, and large penalty when x_i is out of $[\alpha,\ \beta]$, compared to the Tikhonov regularization. Let $R_{com}(\boldsymbol{x})$ denote the combined l_2-l_p-norm penalty function:

$$R_{com}(\boldsymbol{x}) = \sum_{i=1}^{N} r_{com}(x_i), \tag{3.28}$$

where $r_{com}(x_i)$ is the cost function for the ith fluorophore yield coefficient. $r_{com}(x_i)$ has the following properties:

1. $r_{com}(x_i)$ is continuous and differentiable, $\forall x_i \in \mathbb{R}$.
2. When $|x_i| \leq \alpha$, $r_{com}(x_i) \geq r_2(x_i)$.
3. When $\alpha < |x_i| < \beta$, $r_{com}(x_i) < r_2(x_i)$.
4. When $|x_i| \geq \beta$, $r_{com}(x_i) \geq r_2(x_i)$.

The cost function is given as follows:

$$r_{com}(x) = \begin{cases} c_1 x^2, & |x| \leq \kappa_1; \\ c_2|x|^p + c_3, & \kappa_1 \leq |x| \leq \kappa_2; \\ c_4 x^2 + c_5|x|^p + c_6, & |x| > \kappa_2. \end{cases} \tag{3.29}$$

where κ_1 and κ_2 are two constants satisfying $0 < \kappa_1 < \alpha < \kappa_2 < \beta$. (A.1) could satisfy all the four properties if the coefficients c_1, c_2, c_3, c_4, c_5, c_6 are chosen as follows:

$$c_1 = \frac{p\alpha^2 \kappa_1^{p-2}}{2\alpha^p + (p-2)\kappa_1^p}, \tag{3.30}$$

$$c_2 = \frac{2\alpha^2}{2\alpha^p + (p-2)\kappa_1^p}, \tag{3.31}$$

$$c_3 = \frac{(p-2)\alpha^2 \kappa_1^p}{2\alpha^p + (p-2)\kappa_1^p}, \tag{3.32}$$

$$c_4 = \frac{p\kappa_2^{p-1}(\beta^2 - c_2\beta^p - c_3)}{p\kappa_2^{p-1}\beta^2 + (2-p)\kappa_2^{p+1} - 2\kappa_2\beta^p}, \tag{3.33}$$

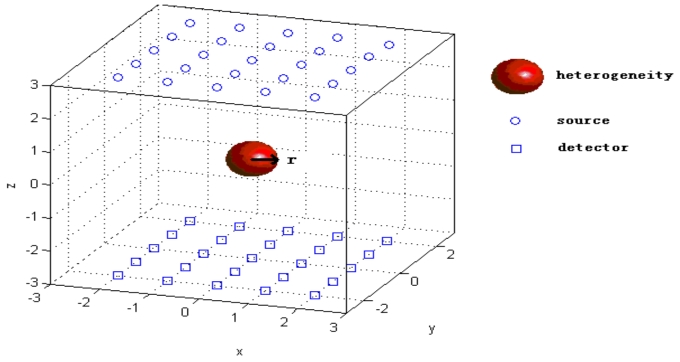
$$c_5 = c_2 - \frac{2\kappa_2^{2-p} c_4}{p}, \tag{3.34}$$

$$c_6 = \beta^2 - \beta^2 c_4 - \beta^p c_5. \tag{3.35}$$

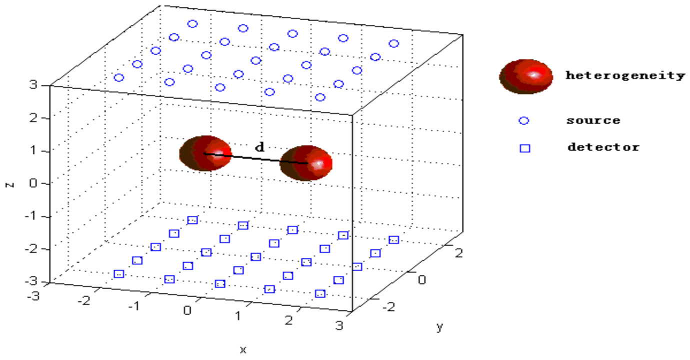
The detailed calculation of the coefficients in the cost function is given in the Appendix A. Since $R(\boldsymbol{x})$ is continuous and differentiable, the solution can be obtained by standard gradient based optimization techniques, such as generalized conjugate gradient or Quasi-Newton method [44,120]. Figure 3.2 shows a comparison of $r_{com}(x)$ with the l_2-norm constraint, when $p = 1$, $\kappa_1 = 0.1$, $\alpha = 1, \kappa_2 = 1.5, \beta = 2$. It's clear that for $|x| \leq \alpha$ or $|x| \geq \beta$, $r_{com}(x_i) \geq r_2(x_i)$, else $r_{com}(x_i) < r_2(x_i)$.

3.4 Numerical Simulations

In this section, numerical simulations results are presented to compare the performances of different regularization techniques for FDOT inverse problem. The regularization techniques we used here include: SVD, CG, ART, the Tikhonov regularization, the l_p-norm regularization ($p = 1$ and $p = 1.2$) and the combined



(a) Single heterogeneity



(b) Double heterogeneity

Figure 3.3: Digital phantom simulation setup.

$l_2 - l_1$-norm regularization. Both visual images and quantitative measurements are used to compare the reconstruction results of different regularization techniques.

3.4.1 Simulation Setup

The simulated imaging domain is a 6 cm × 6 cm × 6 cm cubic region with 25 sources (marked as circles) and 25 detectors (marked as squares) uniformly distributed on the top and bottom surfaces, respectively, as shown in Figure 3.3. The spacing between adjacent sources (detectors) is 1 cm. We first reconstructed the fluorophore concentration of the phantom with single heterogeneity, as shown in Figure 3.3(a). The fluorophore heterogeneity has 3 different radii, $r = 0.25, 0.5, 0.75$ cm, respectively. Next, we reconstructed the fluorophore concentration of the phantom with two heterogeneities, both have radius $r = 0.5$ cm, and the distances between the centers of the heterogeneity are $d = 1.0, 1.5, 2.0$ cm respectively.

We assume homogeneous diffusion coefficient and endogenous absorption coefficient over the imaging domain, $\mu'_{s,x} = \mu'_{s,m} = 10$ cm^{-1} ($D = 1/3(\mu_a + \mu'_s)$), and $\mu_{ae} = 0.05$ cm^{-1}; the fluorophore absorption coefficient inside the spheres are $\mu_{af} = 0.005$ cm^{-1}. We define the *signal-to-noise ratio* (SNR) as follows:

$$SNR = 20 \log_{10} \frac{\|\mathbf{\Gamma}\|_2}{\|\boldsymbol{\epsilon}\|_2}. \tag{3.36}$$

Four different noise levels were added to the measurement, with SNR$= 20, 30, 40, 50$ dB respectively.

3.4.2 Singular Value Spectrum of the Forward Matrix

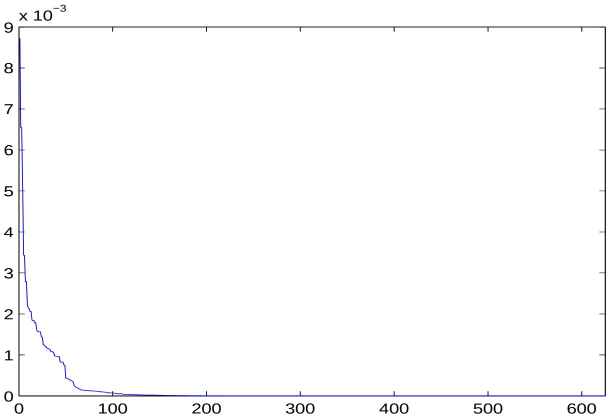
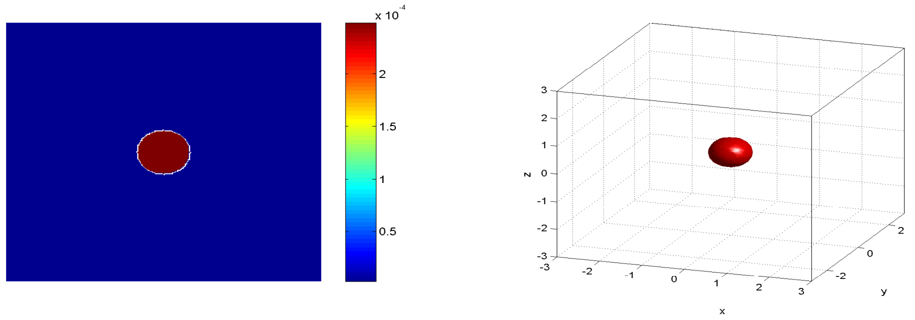


Figure 3.4: Singular value spectrum of the forward matrix.

We calculated the forward matrix $\boldsymbol{A}$ using the finite element method mentioned in Chapter 2. Figure 3.4 shows the singular value spectrum of the forward matrix $\boldsymbol{A}$ in the digital phantom simulation. The forward matrix is ill-posed with a large range of singular values very close to 0, which indicates the necessity of regularization techniques for the reconstruction.

3.4.3 Visual Reconstruction Results

3.4.3.1 Phantom with Single Heterogeneity



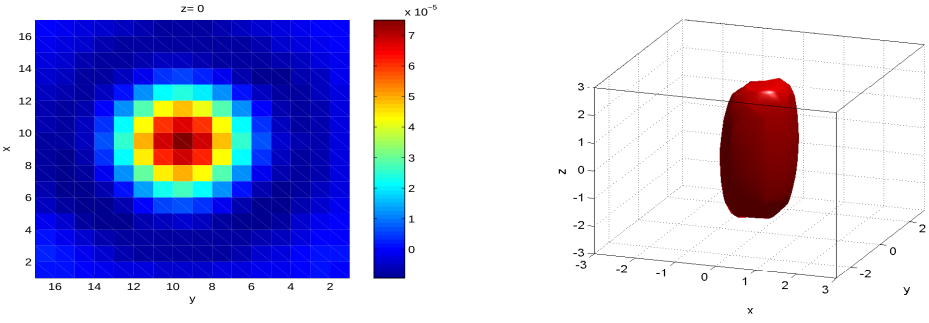
(a) Cross sectional plane at $z = 0$ (b) 3D surface of the heterogeneity

Figure 3.5: Fluorophore configuration of the Phantom with single heterogeneity of radius 0.5 cm.

In this section, we show some examples of the visual reconstruction results using different regularization techniques. Figure 3.5 shows the true fluorophore configuration of the phantom with one heterogeneity of 0.5 cm radius. Figure 3.5(a) shows a 2D cross-sectional image of the phantom at the center ($z = 0$) plane, and Figure 3.5(b) shows the surface of the spherical heterogeneity in 3D.

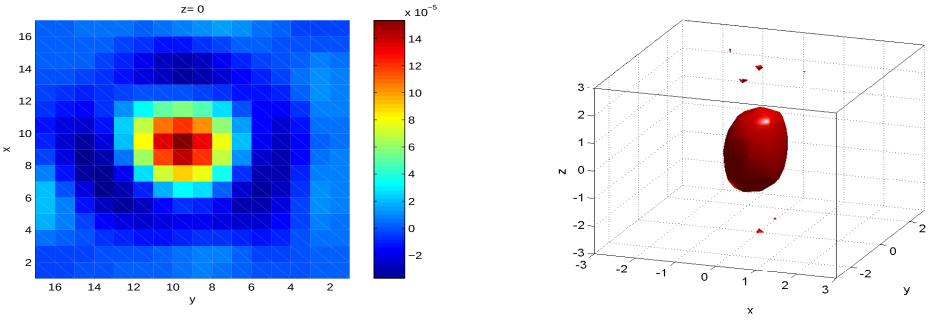
Figure 3.6-3.12 show the reconstructed images of the phantom with 50 dB noise using different regularization techniques mentioned above respectively. From all of the reconstructed images using different regularization techniques, we were able to detect a single heterogeneity at the center of the phantom. However, compared with the true fluorophore configuration in Figure 3.5, sizes of the reconstructed spherical heterogeneities from TSVD, CG, ART and the Tikhonov regularization are overestimated, while the amplitude of the fluorophore coefficients are underestimated as shown in Figure 3.6-3.9. These results are due to the over-smoothing effect in the regularizations. On the other side, the size of the reconstructed spherical heterogeneity by the l_p ($p = 1$ and $p = 1.2$) norm regularization is underestimated and the amplitudes of the fluorophore coefficients at some voxels are overestimated. Because the l_p ($p = 1$ and $p = 1.2$) norm regularization tend to preserve large values in the solution and smooth out small values, the resultant solution has only a few fluorophore yield coefficients above 50% of the maximum amplitude, as shown in

Figure 3.10 and Figure 3.11. Finally, the size of the reconstructed fluorophore heterogeneity by the combined l_2-l_p-norm regularization look closest to the true size. There is an overestimation of the fluorophore yield coefficients in some voxels, however, the overestimation is small compared with the l_p ($p = 1$ and $p = 1.2$) norm regularization, as shown in Figure 3.12.



(a) (b)

Figure 3.6: The SVD reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

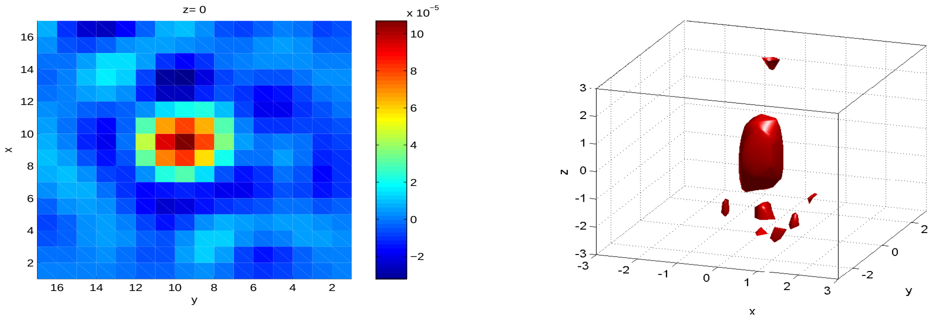


(a) (b)

Figure 3.7: The CG reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

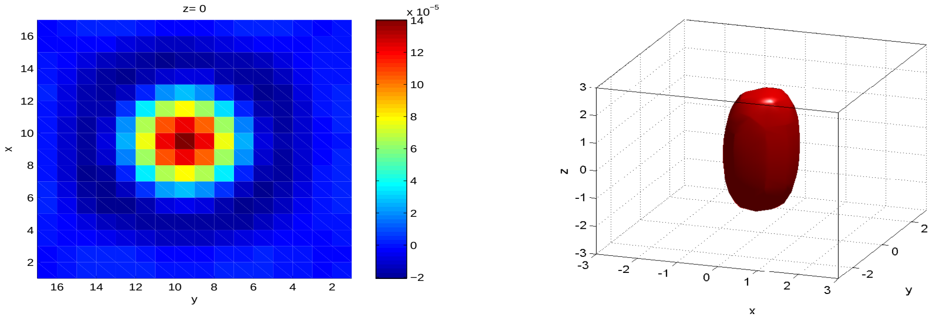
3.4.3.2 Phantom with Two Heterogeneities

Next, we reconstructed the fluorophore yield coefficients in the second-type phantoms which have two spherical heterogeneities each of radius 0.5 cm. Figure 3.13 shows the true fluorophore configuration of the phantom with the centers of the



(a) (b)

Figure 3.8: The ART reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

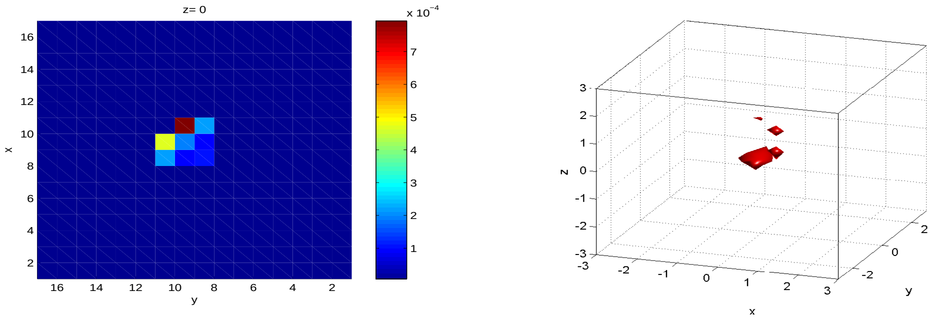


(a) (b)

Figure 3.9: The Tikhonov reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

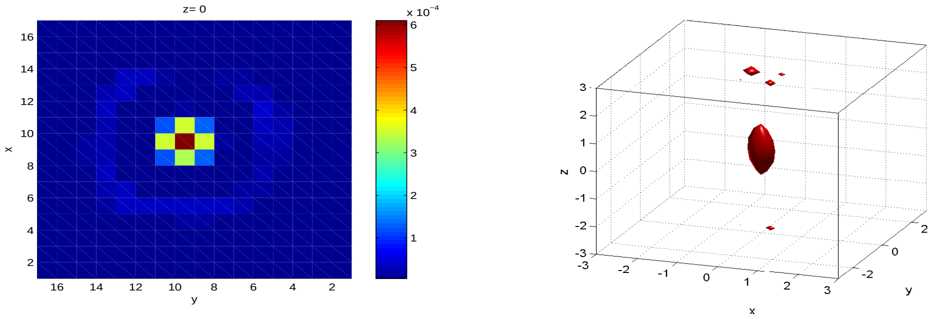
two spherical heterogeneities 2 cm apart. Figure 3.13(a) shows a 2D cross-sectional image of the phantom at $z = 0$ plane, and Figure 3.13(b) shows the surface of the spherical heterogeneity in 3D.

Figure 3.14-3.20 show the reconstructed images of the phantom with 50 dB additive noise using different regularization techniques, respectively. We noticed that the SVD reconstruction in Figure 3.14 results in an over-smoothed image, where the two heterogeneities are merged together and can hardly be identified clearly. In the ART reconstruction, the two heterogeneities are merged, but are still identifiable, as shown in Figure 3.16. In the reconstruction results of CG and the Tikhonov regularization, the two heterogeneities are separated and can be identified as shown



(a) (b)

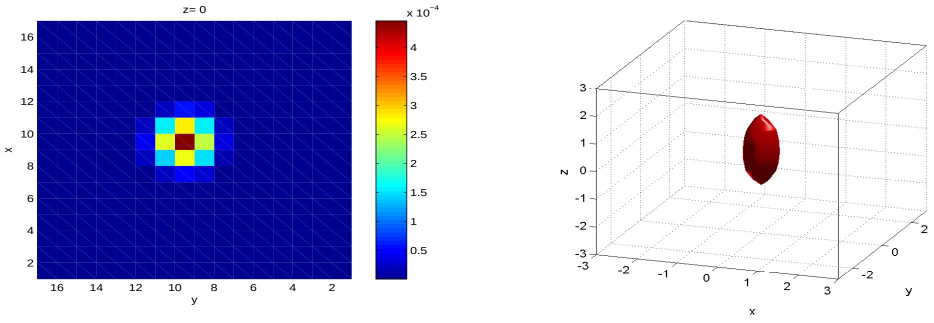
Figure 3.10: The l_1-norm reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



(a) (b)

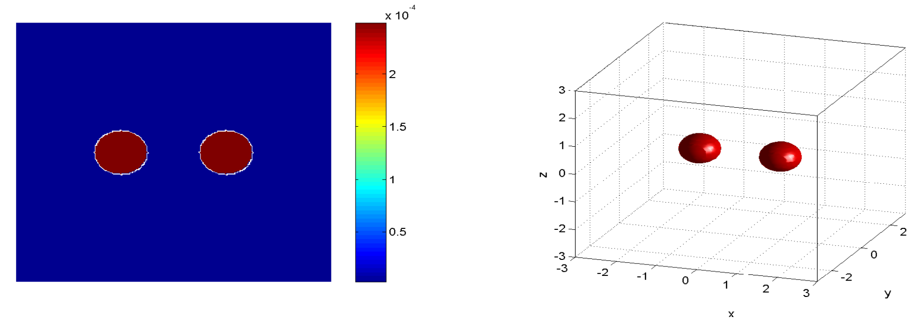
Figure 3.11: The l_p-norm ($p = 1.2$) reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

in Figure 3.15 and Figure 3.17. However, the images look over-smoothing, where the sizes of the heterogeneities look overestimated, and the amplitudes of the reconstructed fluorophore yield coefficients are underestimated. In the l_p ($p = 1$ and $p = 1.2$) norm reconstruction, the sizes of the reconstructed heterogeneities look smaller than the true size, and the fluorophore coefficients at some voxels are overestimated. Compared with the true fluorophore yield map shown in Figure 3.13, we noticed that the reconstruction results by the combined l_2-l_p-norm regularization best approximate the true fluorophore yield map, in the sense that the two heterogeneities are clearly separated, the sizes and the amplitudes of the fluorophore



(a) (b)

Figure 3.12: The combined l_2-l_p-norm ($p = 1$) reconstruction results ($r = 0.5$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



(a) Cross sectional plane at $z = 0$ (b) 3D surface of the heterogeneity

Figure 3.13: Fluorophore configuration of phantom with two heterogeneities of radius 0.5 cm (2 cm apart in the center).

heterogeneities are close to the true case.

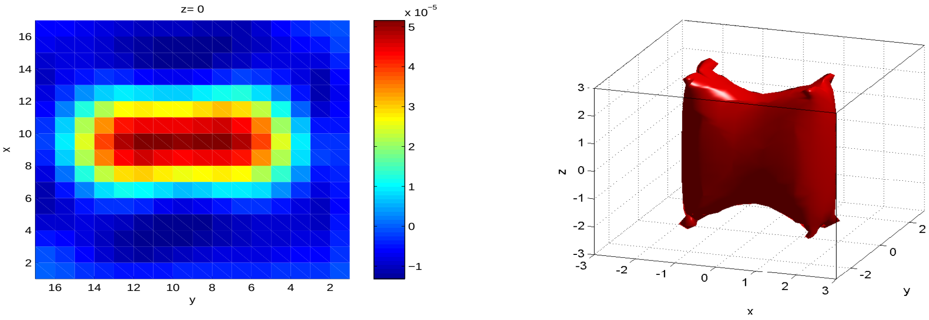
3.4.4 Quantitative Performance Measurements

In this section, we used several quantitative measurements to evaluate different regularization techniques.

We first evaluated the *mean square error* (MSE) between the real fluorophore yield and the reconstructed image over the entire image domain,

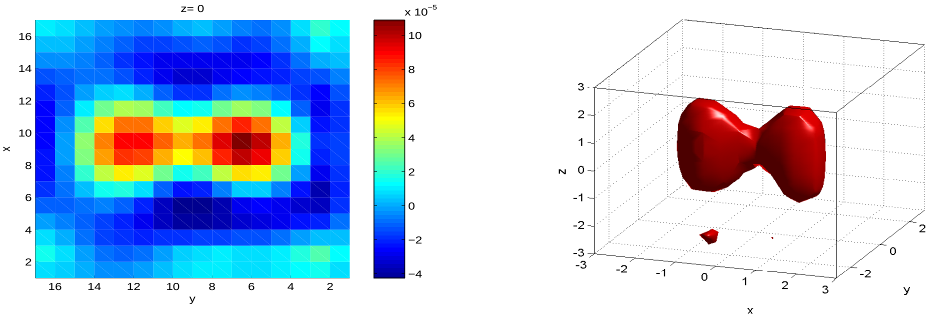
$$\text{MSE} = \frac{\|\boldsymbol{x}_{\text{true}} - \boldsymbol{x}_{\text{recon}}\|_2}{\#\text{ of voxels}}. \tag{3.37}$$

Figure 3.21 shows the MSE versus SNR for all the six different phantom



(a) (b)

Figure 3.14: The SVD reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

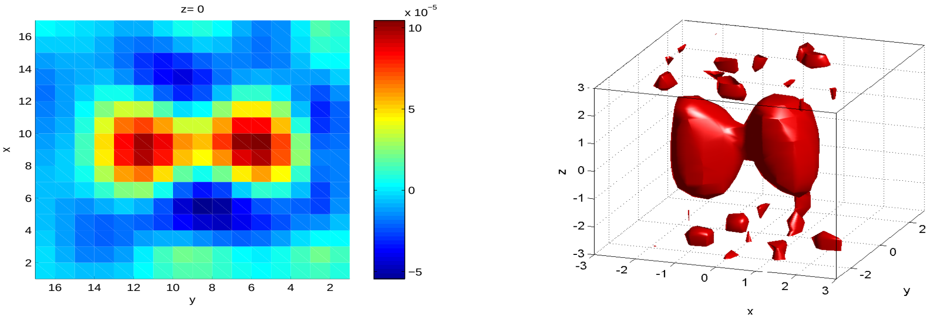


(a) (b)

Figure 3.15: The CG reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

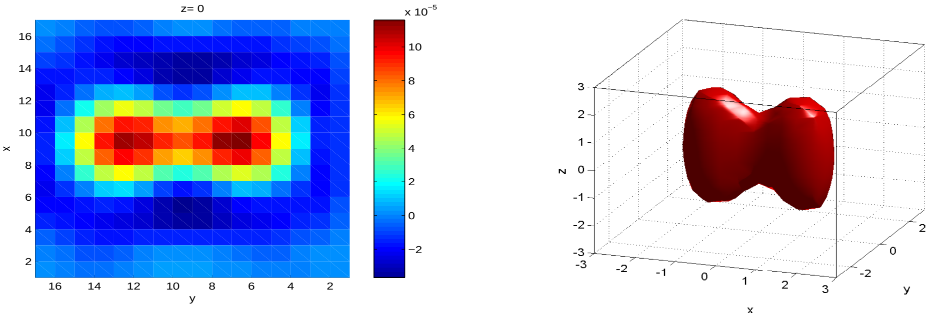
configurations. The MSE in ART is the largest in all of the six phantoms when SNR<40dB, which indicates that ART reconstruction is sensitive to the noise.

The second performance measure is the *contrast-to-object noise ratio* (CONR), defined as the ratio of the mean value of the image in the true fluorophore region to the mean image standard deviation in the same region [109]. The true fluorophore yield map is piece wise constant, where the standard deviation in the foreground region is close to zero, thus large CONR is expected. Figure 3.22 shows the CONR versus SNR for all of the six phantoms we tested respectively. The Tikhonov regularization and the TSVD have the largest CONR, which indicates that the recon-



(a) (b)

Figure 3.16: The ART reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

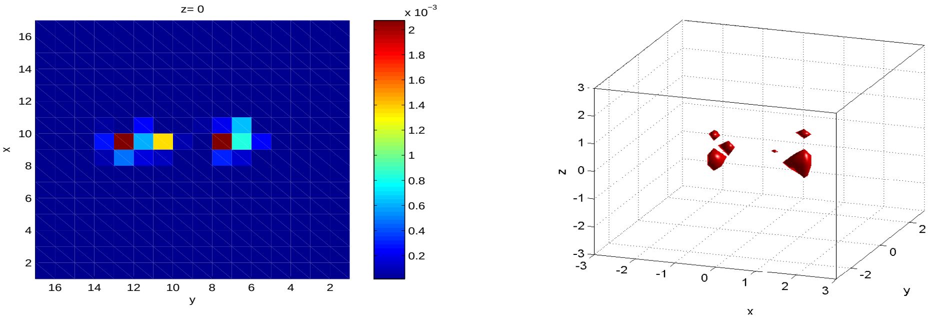


(a) (b)

Figure 3.17: The Tikhonov reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

structed images are smooth in the foreground region. On the other side, the l_p-norm ($p = 1$ and $p = 1.2$) regularization has small CONR, which indicates large standard deviations in the foreground region. The reconstructed images by the combined l_2-l_p-norm regularization, ART and CG have much larger CONR than the l_p-norm ($p = 1$ and $p = 1.2$) regularization, but are less smooth compared with the Tikhonov regularization and TSVD.

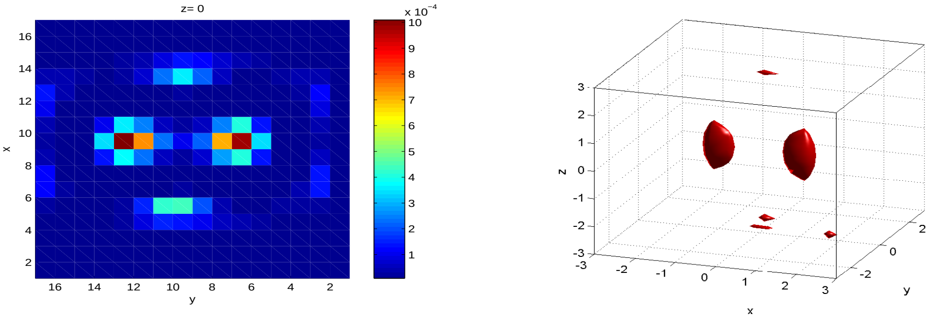
The third performance measure is the *contrast-to-background noise ratio* (CBNR), defined as the ratio of the mean value of the image in the true fluorophore region to the mean image standard deviation in the background region [109]. The true fluo-



(a)

(b)

Figure 3.18: The l_1-norm reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

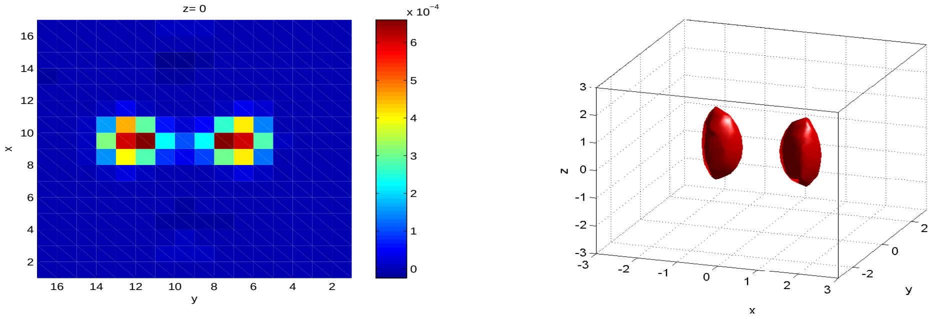


(a)

(b)

Figure 3.19: The l_p-norm ($p = 1.2$) reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

rophore yield map has large mean value in the foreground region, while the standard deviation in the background is zero, thus will have large CBNR. Generally speaking, the larger CBNR indicates better fluorophore contrast between the foreground and background. Figure 3.23 shows the CBNR versus SNR for all the six phantoms we tested respectively. The l_p-norm ($p = 1$ and $p = 1.2$) and the combined l_2-l_p-norm regularization have higher CBNR than other regularization method. Specifically, at high SNR level, the l_1-norm regularization has the largest CBNR value, and the combined l_2-l_p-norm regularization is suboptimal. This is because the reconstructed image by the l_1-norm regularization tends to preserve large spikes in the recon-



(a) (b)

Figure 3.20: The combined l_2-l_p-norm ($p = 1$) reconstruction results (two heterogeneities, $r = 0.5$ cm, $d = 2$ cm, 50 dB additive noise) of (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.

struction, which directly increases the CBNR. However, as the SNR level decreases, the combine l_2-l_p-norm regularization has higher CBNR than the l_1-norm regularization, which indicates that the combined l_2-l_p has lower background standard deviation compared with the l_1-norm regularization. We noticed that although the Tikhonov regularization and TSVD have the largest CONR, they have rather poor fluorophore contrast (much smaller CBNR) compared with the combined l_2-l_p-norm regularization.

The fourth performance measure we used is the *center error* (CE), which is defined as the distance between the center of the true fluorophore concentration and that of the reconstructed fluorophore. We calculated the CE for the phantom configurations with a single heterogeneity. To find the center of the reconstructed fluorophore map, we first found the voxel of the maximum amplitude and classified it as the heterogeneity. Next, we iteratively increased the heterogeneity size by including the neighboring voxels which have amplitudes above 50% of the maximum amplitude value. Once we had classified all the heterogeneity voxels, we computed the centroid as the weighted average of the position of the object voxels, and used the amplitude of these voxels as the weights. Figure 3.24 shows the CE the reconstructed images. As the size of the heterogeneity increases, the overall level of CE also increases. In addition, we noticed that the CE for ART is large at low SNR level, which indicates that ART is sensitive to the increasing noise. This result is consistent with the result in the MSE measurement.

The last performance measure is to evaluate the *spacial resolution* (SR) when multiple heterogeneities are presented in the imaging domain. From the visual results, we classified the reconstructed images into 4 cases:

Case A. The two heterogeneities are clearly separated, such as shown in Fig. 3.25(a).

Case B. The two heterogeneities are not well-separated but still differentiable as shown in Figure 3.25(b).

Case C. The two heterogeneities are well-separated and can be easily identified, but artifacts in the form of small heterogeneities appear in the reconstructed images as shown in 3.25(c).

Case D. The two heterogeneities cannot be detected because, either (i) the two heterogeneities are totally merged together as shown in Figure 3.25(d); or (ii) more than two smaller sized heterogeneities appear as shown in Figure 3.25(e).

The classification of the reconstruction results of Phantom 4-6 using different regularization techniques are shown in Table 3.1-3.3 respectively, each reconstruction at different noise level is classified as Case A, B, C or D. We noticed that the l_p ($p = 1.2$) norm regularization and the combine l_2-l_p-norm regularization give the best resolution of the object. The l_1-norm regularization is likely to produce a number of small heterogeneities at low SNR level, which is due to the large noise in the reconstruction. The reconstruction images by TSVD CG, ART and the Tikhonov regularization are likely to show two large heterogeneities merged together, which is due to the over-smoothing effect.

Table 3.1: Spatial resolution of the reconstructed images, $r = 0.5$ cm, $d = 1.0$ cm.

SNR-Method	TSVD	CG	ART	Tikhonov	l_1	$l_{1.2}$	Combined
50	D	D	D	D	C	C	B
40	D	D	D	D	D	D	D
30	D	D	D	D	D	D	D
20	D	D	D	D	D	D	D

Table 3.2: Spatial resolution of the reconstructed images, $r = 0.5$ cm, $d = 1.5$ cm.

SNR-Method	TSVD	CG	ART	Tikhonov	l_1	$l_{1.2}$	Combined
50	D	B	D	B	C	A	A
40	D	B	D	B	C	A	A
30	D	D	D	D	D	C	C
20	D	D	D	D	D	D	D

Table 3.3: Spatial resolution of the reconstructed images, $r = 0.5$ cm, $d = 2.0$ cm.

SNR-Method	TSVD	CG	ART	Tikhonov	l_1	$l_{1.2}$	Combined
50	D	A	B	A	A	A	A
40	D	B	B	B	C	A	A
30	D	D	D	D	D	C	C
20	D	D	D	D	D	D	D

3.5 Silicon Phantom Experiment

In this section, we show the 3D reconstruction results of a silicon phantom, which has a cylindrical hole in the center filled with a mixture of intralipid, ink and Cy7. The imaging domain is approximated as a 2 cm×2 cm×2 cm rectangular region, and the diameter of the hole in the center of the phantom is approximately 3mm, as shown in Figure 3.26. The optical property for the phantom is : $\mu_a = 0.2$ cm^{-1} and $\mu_s' = 12$ cm^{-1} ($D = 1/3(\mu_a + \mu_s')$). The concentration of Cy7 inside the cylindrical hole is 1μ M and 0 outside the hole. We use 48 sources and 12 detectors on the top and bottom of the phantom respectively.

Figure 3.27-3.33 show the reconstruction results using different regularization techniques respectively. Compared with Figure 3.26, the reconstruction results by TSVD, CG, ART and the Tikhonov regularization show over-smoothed solutions in Figure 3.27-3.30, that the reconstructed tubes are much larger than that in Figure 3.26. While the reconstructed fluorophore tube by the l_1-norm regularization look small in size, as shown in Figure 3.31. The size of the reconstructed fluorophore tube by the $l_{1.2}$-norm regularization is slight larger than that of the l_1-norm regularization. The combined l_2-l_p-norm regularization produces less over-smoothed images compared with TSVD, CG, ART and the Tikhonov regularization, and the size of the reconstructed fluorophore tube looks slightly larger than the true size, as shown in Figure 3.33.

The CONR and CBNR for different regularization methods are given in Table 3.4. The combined l_2-l_p-norm regularization has the largest CBNR, which indicates that it produces the highest image contrast. We notice that the l_p-norm ($p = 1$ and $p = 1.2$) regularization also have large CBNR values, but they have the smallest CONR as well, which indicates less smoothness in the foreground region. The TSVD and the Tikhonov regularization produce smooth images (the largest CONR) in the expense of low image contrast (the smallest CBNR). We notice that although the combined l_2-l_p-norm regularization have slight smaller CONR than the TSVD and the Tikhonov regularization, it gives much larger CBNR. This result is consistent with the results from digital phantom simulation in Section 3.4.

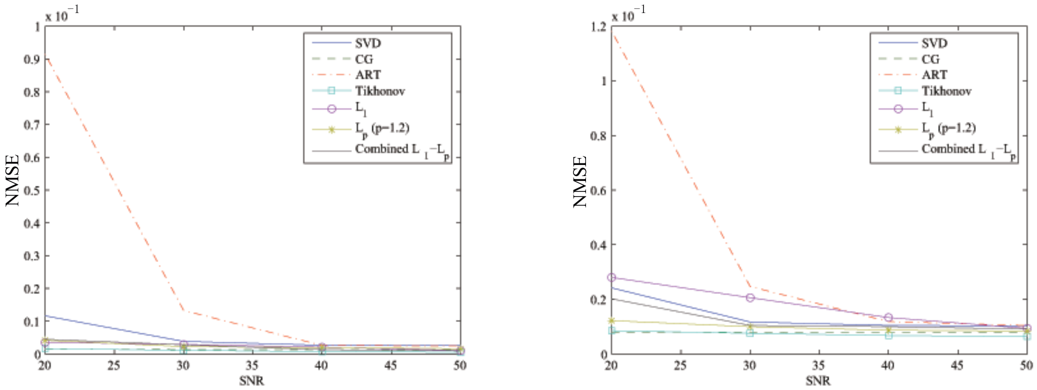
Table 3.4: The CONR and CBNR for different regularization methods.

Method	TSVD	CG	ART	Tikhonov	l_1	$l_{1.2}$	Combined
CONR	4.43	3.35	3.26	4.03	1.08	1.39	3.97
CBNR	4.77	4.65	5.07	6.79	7.98	7.93	8.21

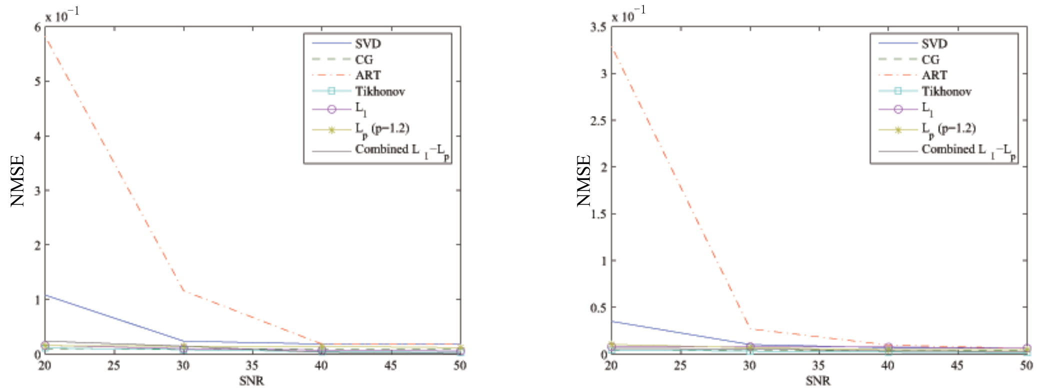
3.6 Conclusion

In this chapter, we proposed a combined l_2-l_p-norm regularization method to reconstruct the fluorophore yield in the FDOT inverse problem. Different from the l_2 norm regularization or the l_1-norm regularization, the proposed method is able to preserve the fluorophore yield coefficients that have large amplitudes in the foreground region, while smooth out large spikes due to the measurement noise or the ill-conditioning of the forward matrix. We tested the combined l_2-l_p-norm regularization using both synthetic data generated by 3D numerical simulation and real silicon phantom experiment. We also compared the proposed regularization method with other commonly used regularization techniques, including TSVD, CG, the Kaczmarz iteration method, the Tikhonov regularization and the l_p norm ($p = 1$ and $p = 1.2$) regularization. A number of quantitative measurements were compared for different regularization techniques.

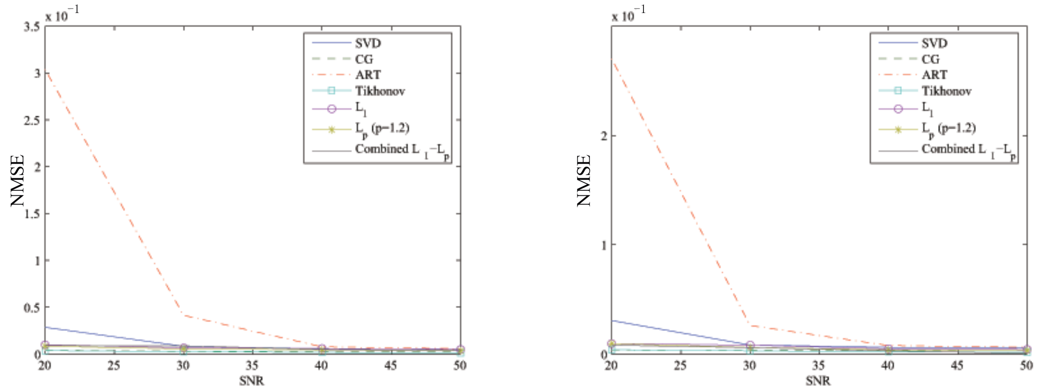
We showed that the proposed method gives the best estimate to the size of the fluorophore heterogeneity and provides the highest visual resolution. It also has low errors in all of the quantitative error measurements, provides high fluorophore contrast between the foreground and background region, and preserves the smoothness of the reconstructed image.



(a) Single heterogeneity, $r = 0.25$ cm (b) Single heterogeneity, $r = 0.5$ cm

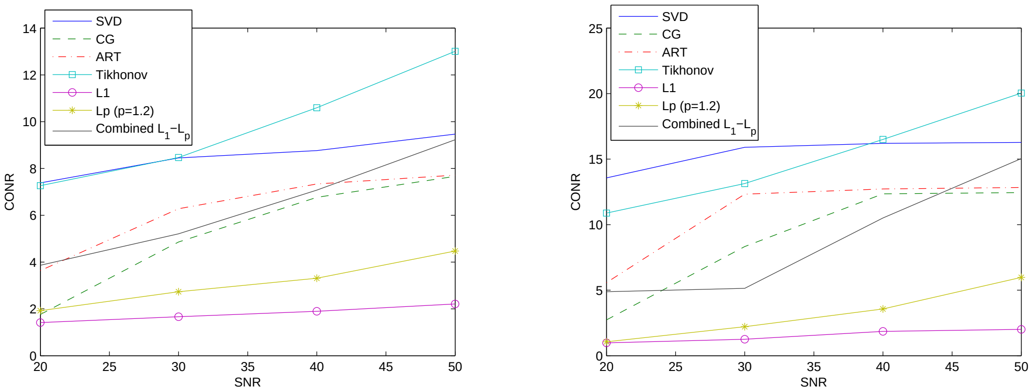


(c) Single heterogeneity, $r = 0.75$ cm (d) Double heterogeneities, $r = 0.5$ cm, $d = 1.0$ cm

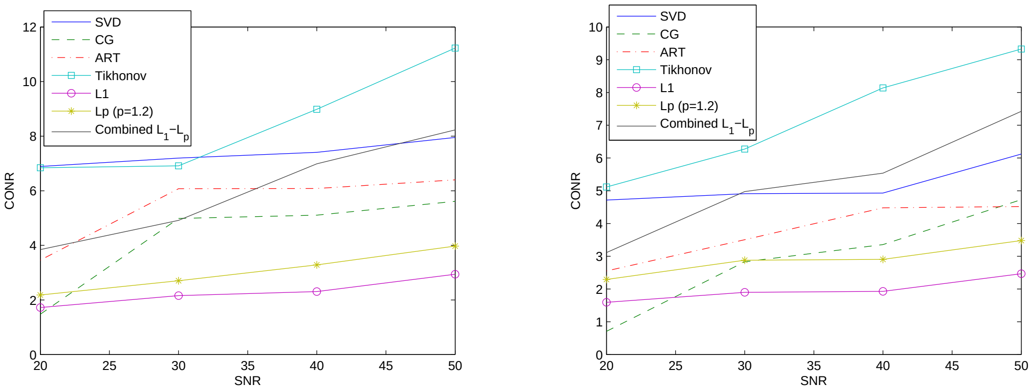


(e) Double heterogeneities, $r = 0.5$ cm, $d = 1.5$ cm (f) Double heterogeneities, $r = 0.5$ cm, $d = 2.0$ cm

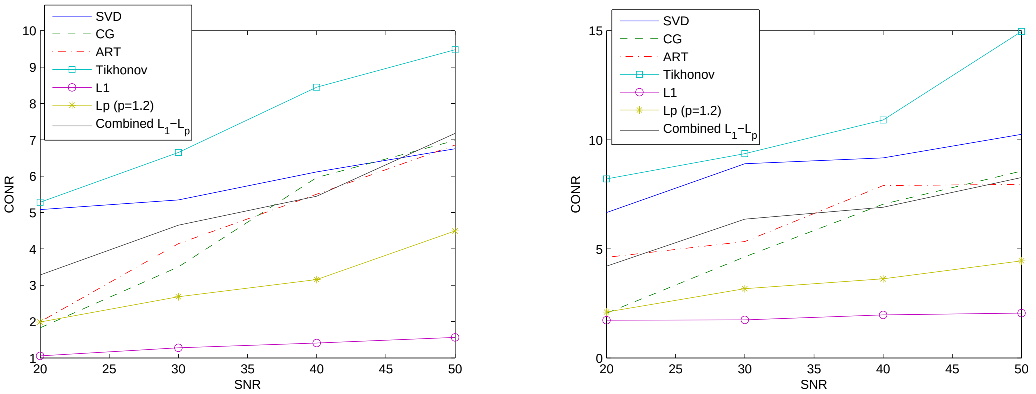
Figure 3.21: MSE-SNR plot of six different phantom configurations.



(a) Single heterogeneity, $r = 0.25$ cm (b) Single heterogeneity, $r = 0.5$ cm

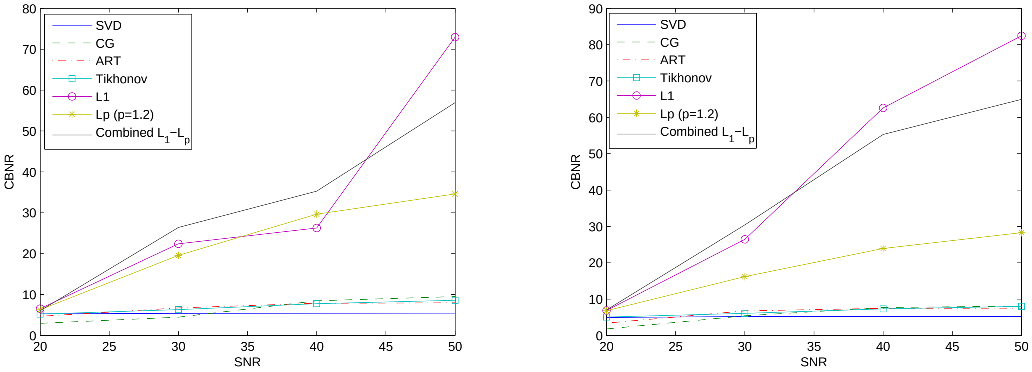


(c) Single heterogeneity, $r = 0.75 cm$ (d) Double heterogeneities, $r = 0.5$ cm, $d = 1.0$ cm



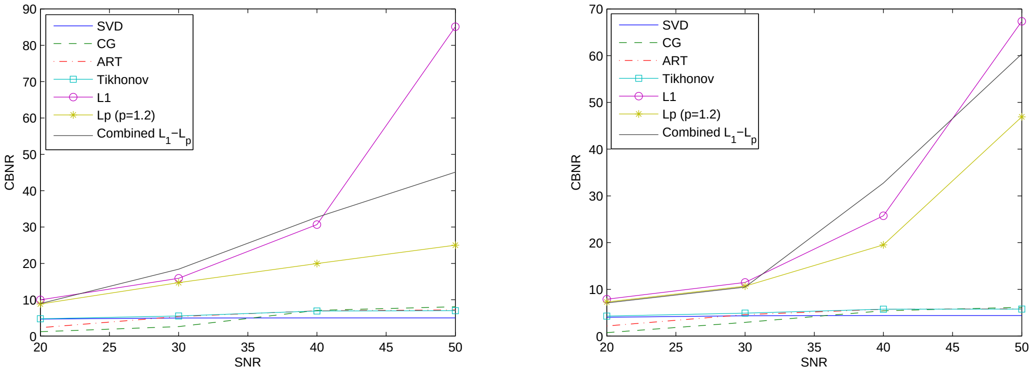
(e) Double heterogeneities, $r = 0.5$ cm, $d = 1.5$ cm (f) Double heterogeneities, $r = 0.5$ cm, $d = 2.0$ cm

Figure 3.22: CONR-SNR plot of six different phantom configurations.



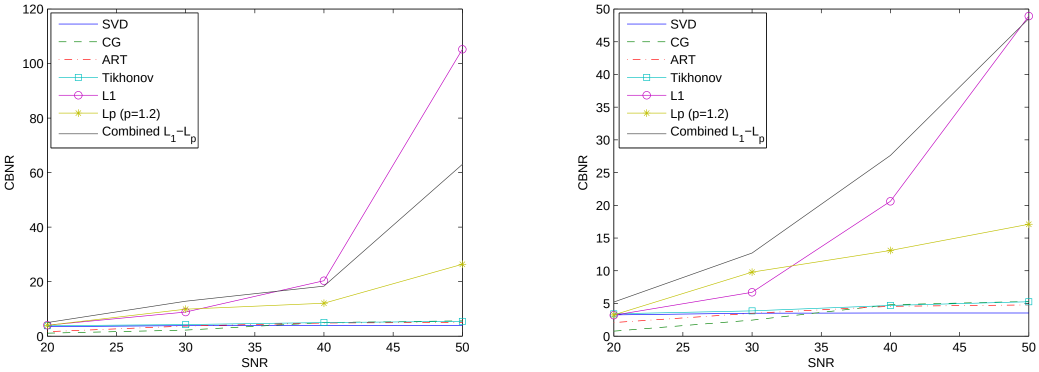
(a) Single heterogeneity, $r = 0.25$ cm

(b) Single heterogeneity, $r = 0.5$ cm



(c) Single heterogeneity, $r = 0.75$ cm

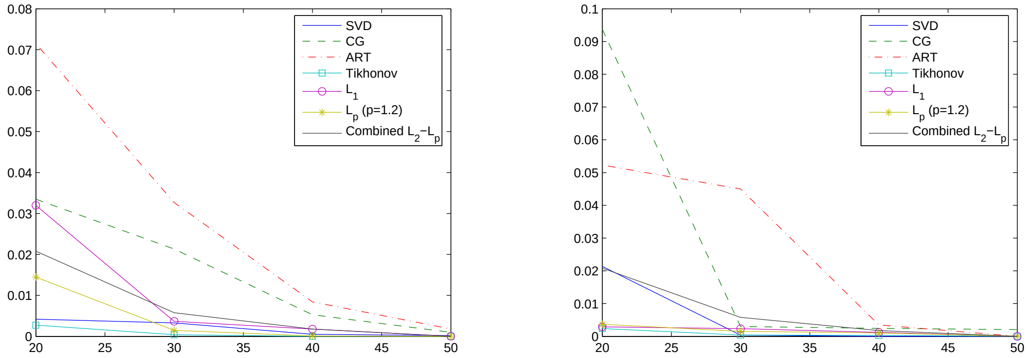
(d) Double heterogeneities, $r = 0.5$ cm, $d = 1.0$ cm



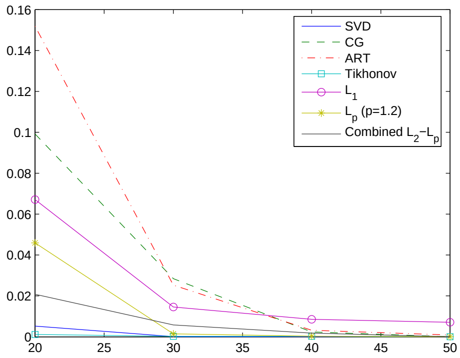
(e) Double heterogeneities, $r = 0.5$ cm, $d = 1.5$ cm

(f) Double heterogeneities, $r = 0.5$ cm, $d = 2.0$ cm

Figure 3.23: CONR-SNR plot of six different phantom 6 configurations.

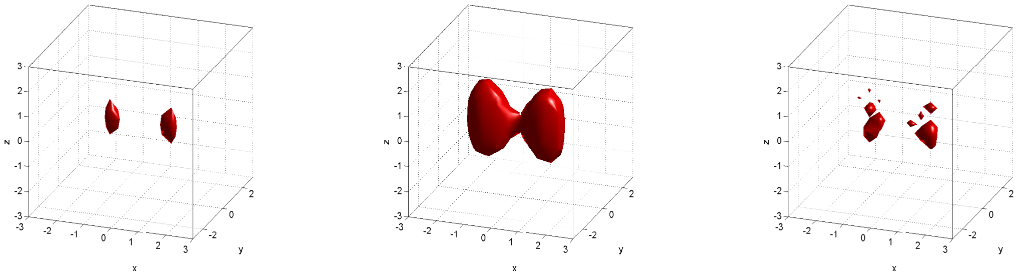


(a) Single heterogeneity, $r = 0.25$ cm (b) Single heterogeneity, $r = 0.5$ cm

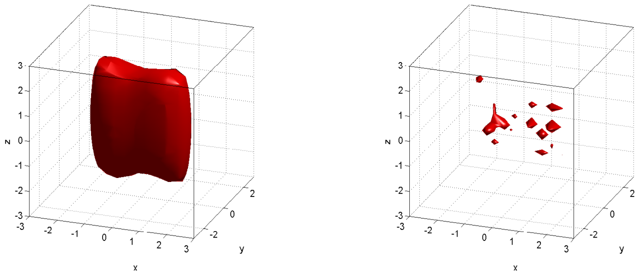


(c) Single heterogeneity, $r = 0.75$ cm

Figure 3.24: CE-SNR plot of different phantom configurations.

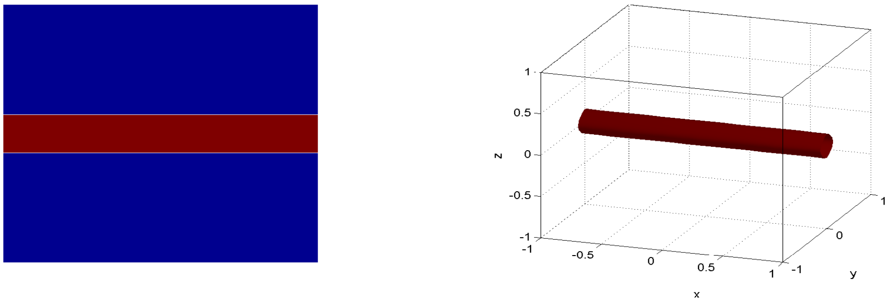


(a) Case A (b) Case B (c) Case C



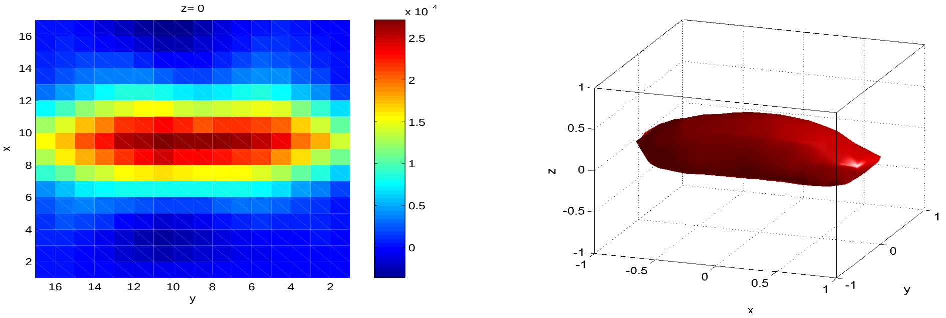
(d) Case D (i) (e) Case D (ii)

Figure 3.25: Examples of the resolutions of heterogeneities.



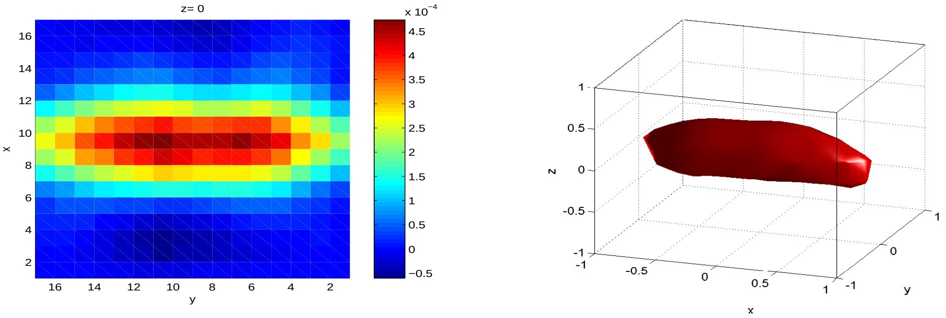
(a) (b)

Figure 3.26: Fluorophore configuration of the real silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D surface of the heterogeneity.



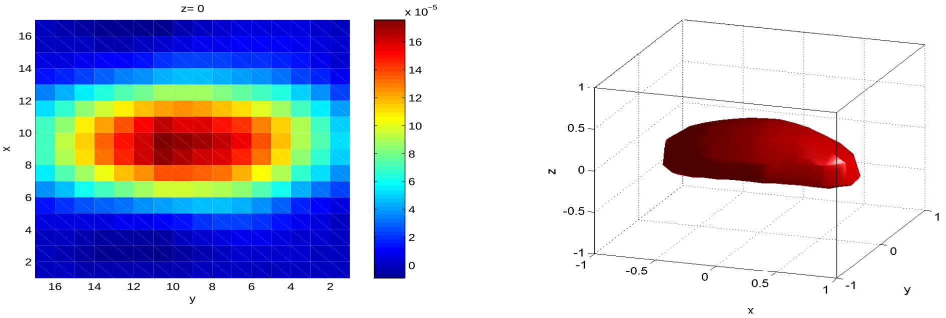
(a) (b)

Figure 3.27: SVD reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



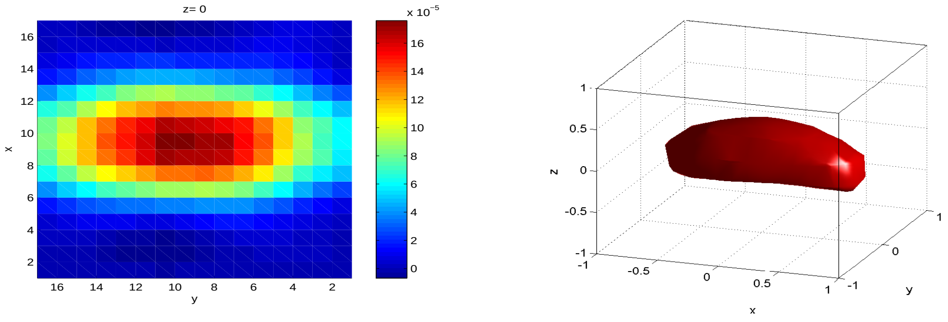
(a) (b)

Figure 3.28: CG reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



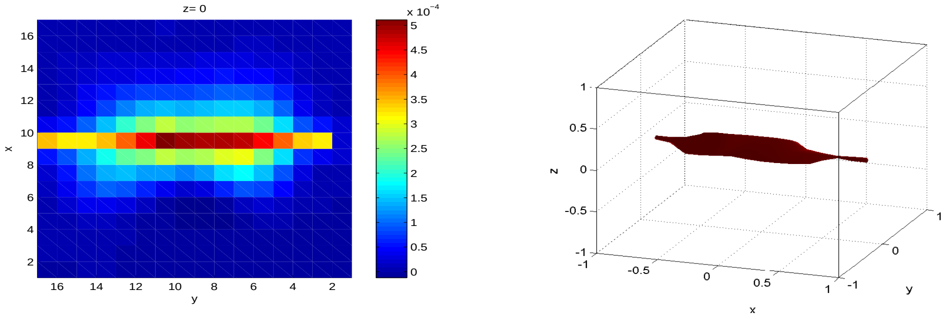
(a) (b)

Figure 3.29: ART reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



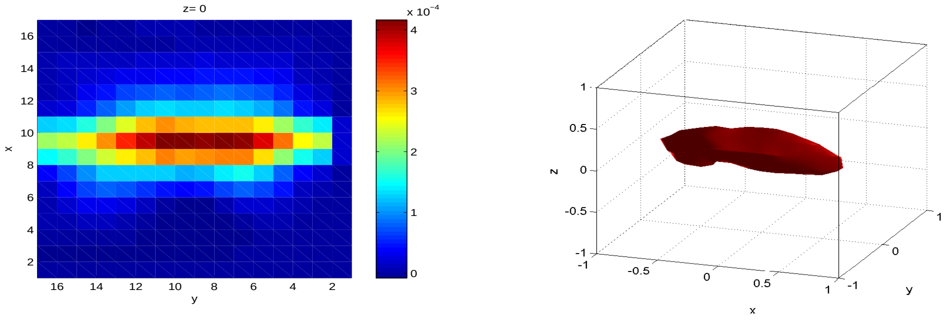
(a) (b)

Figure 3.30: The Tikhonov reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



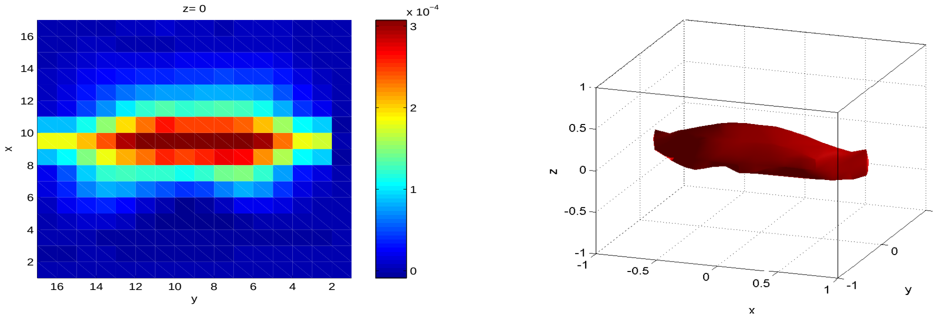
(a) (b)

Figure 3.31: The l_1-norm reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



(a) (b)

Figure 3.32: The l_p $(p = 1.2)$ norm reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D iso-surface at 50% of the maximum amplitude in the reconstruction.



(a) (b)

Figure 3.33: The combined $l_2 - l_p$-norm reconstruction results of silicon phantom: (a) cross sectional plane at $z = 0$; (b) 3D isosurface at 50% of the maximum amplitude in the reconstruction.

CHAPTER 4
SPARSE SIGNAL RECOVERY TECHNIQUES IN FDOT RECONSTRUCTION

4.1 Introduction

In FDOT, the fluorophore concentration typically targets on a small region of interest, e.g., tumors, rather than smoothly distributed over the whole imaging domain [48, 96, 97]. As a result, the fluorophore concentration (or fluorophore yield, which is proportional to the fluorophore concentration) to be reconstructed is often very sparse in the imaging domain. Thus, the FDOT inverse problem can be formulated as recovering sparse signal from an under-determined linear system.

In CS framework, sparse signals, i.e., signals that have a relatively small number of non-zero entries, can be exactly recovered from limited number of measurements [45]. In CS literatures, a wide range of sparsity promoting recovery techniques have been developed [60–65, 67–74], which typically base on greedy algorithms [60–65] or relaxation methods [67–74]. Recently, various sparsity promoting recovery techniques in CS framework has been applied to the inverse problem of optical tomography [48, 48, 95–97], and have shown promising results.

In this chapter, we address the ill-posed FDOT inverse problem by exploiting the sparse nature of the fluorophore yield in the imaging domain. We review a number of sparsity promoting recovery techniques in CS literature. We apply various sparsity promoting recovery techniques to reconstruct the fluorophore yield in the FDOT inverse problem. In the 3D numerical simulation and real phantom experiment, we show that compared with the traditional Tikhonov regularization, most sparsity promoting reconstruction techniques have better reconstruction performance both visually and quantitatively. In addition, we show that in general, relaxation techniques outperforms greedy algorithms in FDOT reconstruction.

The rest of this chapter is organized as follows: Section 4.2 describes the sparse representation of signal; Section 4.3 introduces the sparsity constraint into the FDOT inverse problem formulation; Section 4.4 reviews the basic incoherent sampling property in CS theory; Section 4.5 reviews various sparse signal recovery techniques in the CS literature; In Sections 4.6 and 4.7, we present the performance of different sparse signal recovery techniques in 3D numerical simulations and in a real silicon phantom experiment; Section 4.8 is the discussion; Finally, Section 4.9 concludes the chapter.

4.2 Sparse Representation

The primary interest of compressive sensing is to recover sparse signals, i.e., a large portion of the signal is zero or negligible. Many signals are either naturally sparse or they are sparse when they are projected onto some basis.

Let $\boldsymbol{\theta} \in \mathbb{R}^K$ be an K dimensional signal. Let $\boldsymbol{d}_i \in \mathbb{R}^K,\ (i = 1, ..., K)$, be K orthogonal basis, such that

$$\boldsymbol{\theta} = \sum_{i=1}^{K} x_i \boldsymbol{d}_i = \boldsymbol{D}\boldsymbol{x}, \tag{4.1}$$

where $\boldsymbol{D} = [\boldsymbol{d}_1, ..., \boldsymbol{d}_K] \in \mathbb{R}^{K \times K}$ with its columns being the basis vectors, and $\boldsymbol{x} = [x_1, ..., x_K]^T \in \mathbb{R}^K$ is a vector of coefficients to represent the signal $\boldsymbol{\theta}$ on $\boldsymbol{D}$. $\boldsymbol{D}$ is often called a *dictionary* in CS theory since the vectors $\boldsymbol{d}_i,\ i = 1, ..., K$ are able to represent any signal in $\mathbb{R}^K$ by choosing appropriate coefficients.

We say that the vector $\boldsymbol{x}$ is *sparse* if most of the entries in $\boldsymbol{x}$ are zero or can be neglected. Alternatively, we say that and the signal $\boldsymbol{\theta}$ has a *sparse representation* if $\boldsymbol{x}$ is sparse. The the number of non-zero or significant entries in $\boldsymbol{x}$ is defined as the sparsity of $\boldsymbol{x}$. If the *sparsity* of $\boldsymbol{x}$ is $S\,(S \ll K)$, $\boldsymbol{x}$ is called an S-sparse vector, i.e.,

$$S = \|\boldsymbol{x}\|_0, \tag{4.2}$$

where $\|\cdot\|_0$ denotes the l_0-norm of a vector, which counts the number of non-zero entries. Similarly, we denote the l_p-norm of a vector by $\|\cdot\|_p$, i.e.,

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{K} |x_i|^p \right)^{\frac{1}{p}}. \tag{4.3}$$

In some cases, the signal can be sparsely represented more effectively by an overcomplete dictionary. An *overcomplete dictionary* is defined as $\boldsymbol{D} = [\boldsymbol{d}_1, ..., \boldsymbol{d}_N] \in \mathbb{R}^{K \times N}$ $(K < N)$, where the columns of $\boldsymbol{D}$ span the whole $\mathbb{R}^K$ space. Since the number of basis in $\boldsymbol{D}$ is larger than the dimension of the space, the basis in an overcomplete dictionary can not be all orthogonal to each other. Any signal in $\mathbb{R}^K$ can be represented by a subset of the basis in $\boldsymbol{D}$.

4.3 Signal Recovery with Sparsity Constraint

Recall from Chapter 2, recovering the sparse signal $\boldsymbol{x}$ from

$$\boldsymbol{\Gamma} = \boldsymbol{A}\boldsymbol{x} \tag{4.4}$$

involves solving the following optimization problem,

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad \text{such that} \quad \boldsymbol{\Gamma} = \boldsymbol{A}\boldsymbol{x}, \tag{4.5}$$

where $\|\boldsymbol{x}\|_0$ measures the sparsity of $\boldsymbol{x}$. In the noisy setting

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{4.6}$$

(4.5) is replaced by

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_0 \ \text{ such that } \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \varepsilon. \tag{4.7}$$

4.4 Incoherent Sampling

CS offers a sampling technique for sparse signals. An important concept associated with the sensing matrix $\boldsymbol{A}$ is the *mutual coherence*, which is defined as follows [51, 75]:

$$\mathcal{M}(\boldsymbol{A}) = \max_{p,q,p \neq q} \frac{|\boldsymbol{a}_p^T \boldsymbol{a}_q|}{\|\boldsymbol{a}_p\|_2 \|\boldsymbol{a}_q\|_2}, \tag{4.8}$$

where $\boldsymbol{a}_p$ and $\boldsymbol{a}_q$ are two different columns in $\boldsymbol{A}$. Clearly, $0 \leq \mathcal{M}(\boldsymbol{A}) \leq 1$. Note that the mutual coherence is the largest normalized inner product of two different columns. We say that the two column vectors $\boldsymbol{a}_p$ and $\boldsymbol{a}_q$ are *incoherent* if they are orthogonal or close to being orthogonal, i.e., the absolute value of their normalized inner product is close to zero.

Mutual coherence is used as a measure of the incoherence of the columns in the matrix $\boldsymbol{A}$. An orthogonal matrix has mutual coherence equals to 0. On the other hand, if two columns in $\boldsymbol{A}$ are parallel to each other, the mutual coherence is 1. In CS theory, exact reconstruction of a sparse signal generally requires that the mutual coherence of the sensing matrix to be small, i.e., the columns of $\boldsymbol{A}$ are as orthogonal as possible.

Suppose that $\boldsymbol{x}_0$ is a solution to (4.4), and $\boldsymbol{x}_0$ satisfies

$$\|\boldsymbol{x}_0\|_0 \leq \frac{1}{2}(1 + \frac{1}{\mathcal{M}(\boldsymbol{A})}). \tag{4.9}$$

Then, $\boldsymbol{x}_0$ is necessarily the unique solution of (4.4) with smallest number of nonzero entries [51, 75]. In other words, the linear system (4.4) can mostly have one solution whose sparsity is less than or equal to $\frac{1}{2}(1 + \frac{1}{\mathcal{M}(\boldsymbol{A})})$. Mathematically, $\boldsymbol{x}_0$ is the solution of the following optimization problem in (4.5). In practice, for the noise measurement $\boldsymbol{y}$, there is also a substantial body of work in CS focusing on the noisy setting [55, 121–123]. In case $\|\boldsymbol{\epsilon}\|_2 < \varepsilon$, it has been proved that if the quantity in (5.8) is satisfied, the difference between the solution to (4.7) and real signal to be reconstructed is within a constant multiple of the noise level [55].

The mutual coherence of $\boldsymbol{A}$ is determined by the two least "incoherent" columns, which shows the worst-case analysis of the orthogonality of the columns of $\boldsymbol{A}$. While it provides a strong theoretical guarantee for the unique recovery of sparse signals, it is not a necessary condition for the sparse signal recovery. It was observed that conditions based on some "average" measure of the orthogonality of the forward matrix is more likely to describe the typical performance of the sparse signal recovery methods in a large number of cases [52–54, 66, 77, 78, 124–126]. A number of alternative measures to quantify the "average" coherence of a sensing matrix has been proposed [54, 77, 78]. One of these is the *cumulative coherence* or the *Babel* *function*, which is defined as [77]

$$\mathcal{M}_1(k, \boldsymbol{A}) = \max_{p} \max_{|Q|=k, p \notin Q} \sum_{q \in Q} \frac{| < \boldsymbol{a}_p, \boldsymbol{a}_q > |}{\|\boldsymbol{a}_p\|_2 \|\boldsymbol{a}_q\|_2}, \tag{4.10}$$

where Q is a subset of the columns in $\boldsymbol{A}$. It was shown that, if

$$\mathcal{M}_1(k, \boldsymbol{A}) + \mathcal{M}_1(k+1, \boldsymbol{A}) < 1, \tag{4.11}$$

then, the signal $\boldsymbol{x}_0$ with $\|\boldsymbol{x}_0\|_0 \leq k$ can be exactly recovered using greedy algorithms or basis pursuit [77]. Note that $\mathcal{M}_1(k, \boldsymbol{A})$ is an monotonically non-decreasing function in k, and $\mathcal{M}_1(1, \boldsymbol{A}) = \mathcal{M}(\boldsymbol{A})$. If $\mathcal{M}_1(k, \boldsymbol{A})$ increase slowly, $\boldsymbol{A}$ in said to be quasi-incoherent.

Another important concept in CS related to the average coherence is the *restricted isometry property* (RIP) [127, 128] . A matrix $\boldsymbol{A}$ satisfies RIP of order S if there exists a restricted isometric constant $\delta_S \in (0, 1)$, such that for every submatrix $\boldsymbol{A}_S$ which is composed of any S columns of $\boldsymbol{A}$ satisfies

$$(1 - \delta_S)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}_S\boldsymbol{x}\|_2^2 \leq (1 + \delta_S)\|\boldsymbol{x}\|_2^2, \tag{4.12}$$

where $\boldsymbol{x}$ is an arbitrary S-sparse vector. If (4.12) is satisfied, any S sparse signal can be exactly recovered from (4.5) and (4.31) (or (4.7) and (4.32) in the noisy setting) [127, 128]. RIP focuses on every S columns of the sensing matrix, which reflects the average incoherence of S columns in $\boldsymbol{\Psi}$. The restricted isometric constant δ_S is related to the mutual coherence, which is bounded by $\delta_S \leq (S-1)\mathcal{M}(\boldsymbol{A})$.

4.5 Sparse Signal Recovery Techniques in CS Literature

To look for sparse solutions of under-determined linear systems, commonly used sparse signal recovery strategies are typically based on greedy search strategies [60–65], convex relaxation [67–73], and non-convex relaxation [74]. This section briefly reviews sparsity promoting algorithms in the literature.

4.5.1 Greedy Algorithms

4.5.1.1 Orthogonal Matching Pursuit

Orthogonal matching pursuit (OMP) is a heuristic greedy algorithm to find sparse solutions [60]. For a system $\boldsymbol{\Gamma} = \boldsymbol{A}\boldsymbol{x}$, if $\boldsymbol{A}$ is incoherent, and $\boldsymbol{x}$ satisfies (5.8), OMP is guaranteed to find the exact reconstruction. In the presence of small noise, the reconstruction error is within a constant multiple of noise level [55].

OMP iteratively selects columns from $\boldsymbol{A}$ to represent the measurements in a greedy fashion. In each iteration, the measurement $\boldsymbol{y}$ is approximated by a small set of columns $\boldsymbol{A}_T^{(j)}$, where the subscript $T \subset \{1, 2, ...N\}$ denotes the index set of the column indices of $\boldsymbol{A}$, and superscript j denotes the number of iterations. Thus OMP attempts to solve

$$\boldsymbol{y} \approx \boldsymbol{A}_T^{(j)} \boldsymbol{x}_T^{(j)}. \tag{4.13}$$

The coefficients $\boldsymbol{x}_T^{(j)}$ is given as follows:

$$\boldsymbol{x}_T^{(j)} = \text{Arg}\min_{\boldsymbol{x}_T} \|\boldsymbol{y} - \boldsymbol{A}_T^{(j)} \boldsymbol{x}_T\|_2. \tag{4.14}$$

Here $\boldsymbol{x}_T^{(j)}$ is a sub-vector of $\boldsymbol{x}^{(j)}$ only containing indices in T. All the entries in $\boldsymbol{x}_{T^c}^{(j)}$ are zero (where T^c is the complimentary set of T). The residue is given by

$$\boldsymbol{r}^{(j)} = \boldsymbol{y} - \boldsymbol{A}_T^{(j)} \boldsymbol{x}_T^{(j)}. \tag{4.15}$$

In the next iteration, a new column from $\boldsymbol{A}_{T^c}$ that has the largest correlation with $\boldsymbol{r}^{(j)}$ is added to represent the $\boldsymbol{y}$.

$$\begin{aligned} i &= \text{Arg}\max_i (\tilde{\boldsymbol{a}}_i^T \tilde{\boldsymbol{r}}^{(j)})_i, \\ T^{(j+1)} &= T^{(j)} \cup i \end{aligned} \tag{4.16}$$

where $\tilde{\cdot}_i$ denotes the normalized vector.

In OMP, the residue $\boldsymbol{r}^{(j)}$ is always orthogonal to $\boldsymbol{A}_T^{(j)} \boldsymbol{x}_T^{(j)}$ so that no single columns can be selected twice in subsequent iterations. The expected number of iterations equals to the sparsity of $\boldsymbol{x}$. If the S-sparse signal $\boldsymbol{x}$ meets the sparsity requirement in (5.8), the reconstructed $\hat{\boldsymbol{x}}$ from OMP satisfies [55]

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq \frac{\|\boldsymbol{\epsilon}\|_2}{\sqrt{1 - \mathcal{M}(1 - S)}}.$$

The primary computational burden of OMP lies in calculating the correlation of the residual and the columns in $\boldsymbol{A}$, which requires a computational time of $\mathcal{O}(MN)$. Solving the coefficients $\boldsymbol{x}_T$ involves solving a least square problem as in (4.14). If $\boldsymbol{A}$ is any arbitrary matrix, the running time is $\mathcal{O}(S^2M)$. However, when $\boldsymbol{A}$ has low coherence or satisfies RIP, an iterative method such as *conjugate gradient* (CG) reduces the computational time of solving $\boldsymbol{x}_T$ to just $\mathcal{O}(SM)$ [120]. In case $N \geq S^2$, the running time of each iteration in OMP is $\mathcal{O}(MN)$. To recover an S-sparse signal, S iterations are generally needed. Thus, the total computational bound of OMP is $\mathcal{O}(SMN)$.

4.5.1.2 Gradient Pursuit

The *gradient pursuit* (GP) is developed to approximate the OMP algorithm, and is more computationally efficient than OMP [61]. In each iteration, the measurement $\boldsymbol{y}$ is approximated by a small set of columns $\boldsymbol{A}_T^{(j)}$. Same as OMP, the column indices set T is updated by adding a new element at each iteration as in (4.16). However, in instead of adding only one non-zero entry to $\boldsymbol{x}^{(j)}$ at each iteration, the GP algorithm performs a directional update at each iteration,

$$\boldsymbol{x}_T^{(j)} = \boldsymbol{x}_T^{(j-1)} + c^{(j)}\boldsymbol{d}_T^{(j)}, \tag{4.17}$$

where $c^{(j)}$ is the step size, $\boldsymbol{d}^{(j)}$ is the update direction, and $\boldsymbol{d}_T^{(j)}$ is a sub-vector of $\boldsymbol{d}^{(j)}$ containing only those entries of $\boldsymbol{d}^{(j)}$ with indices in T. Three different ways to estimate $\boldsymbol{d}^{(j)}$ are proposed in [61]:

1. *Gradient pursuit* directly calculate the gradient of $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ with respect to $\boldsymbol{x}$ as the update direction,

$$\boldsymbol{d}_T^{(j)} = \boldsymbol{A}_T^{(j)^T}(\boldsymbol{y} - \boldsymbol{A}_T^{(j)}\boldsymbol{x}_T^{(j-1)}). \tag{4.18}$$

The calculation in (4.18) is fast and straight forward, which leads to expedited approximation to OMP.

2. *Conjugate gradient pursuit* is the second directional optimization scheme [120]. The updated direction $\boldsymbol{d}^{(j)}$ is calculated to be conjugate to all the previous directions $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, ..., \boldsymbol{d}^{(j-1)}$ with respect to the Gramm matrix $(\boldsymbol{A}_T^{(j)^T} \boldsymbol{A}_T^{(j)})$, i.e.,
$$\boldsymbol{d}^{(j)} \boldsymbol{A}_T^{(j)^T} \boldsymbol{A}_T^{(j)} \boldsymbol{d}^{(k)} = 0, \quad \forall\, k < j. \tag{4.19}$$
[61] showed that the *conjugate gradient pursuit* leads to a novel implementation of OMP. However, the computational complexity and storage required are higher compared to the *gradient pursuit* algorithm.

3. An *approximately conjugate gradient pursuit* algorithm was also used to calculate $\boldsymbol{d}_T^{(j)}$ in [61]. Instead of calculating a new update direction that is conjugate with all the previous directions, *approximated conjugate gradient pursuit* calculates the update direction that is conjugate with only a limited number of previous directions. The benefit of the *approximate conjugate gradient pursuit* is that it is much easier to calculate and the memory requirements are significantly reduced.

GP requires computation of correlation between the residual and the columns in $\boldsymbol{A}$, which uses a computational time of $\mathcal{O}(MN)$. Three directional update methods in [61] are provided for the GP algorithm, and the computational time for calculating the directional update by these three methods are $\mathcal{O}(S^2M)$. In practice, the running time of the directional update step in each iteration using *conjugate gradient pursuit* is similar as that of solving the least square problem (4.14) in OMP, but the running time of the directional update using *gradient pursuit* and *approximately conjugate pursuit* is much faster [61]. If $S^2 \leq N$, the computational bound in each iteration is $\mathcal{O}(MN)$.

4.5.1.3 Stagewise Orthogonal Matching Pursuit

In large underdetermined linear systems, the number of iterations required by OMP is large. To make the recovery computationally efficient, one can select more than one columns in each iteration. This is the primary motivation of the *stagewise orthogonal matching pursuit* (StOMP) procedure [62].

In StOMP, the initial residual $\boldsymbol{r}^{(0)} = \boldsymbol{y}$, and the initial estimate $\boldsymbol{x} = 0$. In each iteration, the correlation between the columns in $\boldsymbol{A}$ and the redidual is calculated,

$$c_i = \frac{\boldsymbol{a}_i^T}{\|\boldsymbol{a}_i\|_2} \frac{\boldsymbol{r}^{(j-1)}}{\|\boldsymbol{r}^{(j-1)}\|}, \quad \text{for } i = 1, ..., N. \tag{4.20}$$

If c_i is larger than a threshold value $t^{(j)}$, then $\boldsymbol{a}_i$ is selected to add to the set of columns to present the measurement, i.e.,

$$I_s = \{i | c_i > h^{(j)}\}, \tag{4.21}$$

$$T^{(j)} = T^{(j-1)} \cup I_s. \tag{4.22}$$

A new approximation $\boldsymbol{x}^{(j)}$ supported in $T^{(j)}$ is given by

$$\boldsymbol{x}_T^{(j)} = (\boldsymbol{A}_T^T \boldsymbol{A}_T)^{-1} \boldsymbol{A}_T^T \boldsymbol{y}. \tag{4.23}$$

A critical parameter of StOMP is the threshold value $h^{(j)}$ which is calculated in each iteration. It is given by

$$h^{(j)} = c \|\boldsymbol{r}^{(j-1)}\|_2 / \sqrt{M}, \tag{4.24}$$

where c is a constant parameter. [62] suggests that c to be between 2 and 3. The iteration stops after a fixed number of stages. [62] proved that if $\boldsymbol{A}$ is a random matrix with each element generated from the uniform spherical ensemble, a sparse signal can be exactly recovered in noiseless setting, and is stable under Gaussian noise.

in $\boldsymbol{A}$, and solving least squares systems in (4.23). Unlike OMP, StOMP adds more than one entries into the set T in each iteration as in (4.21)-(4.22), which enables fast implementation. This feature makes StOMP suitable for large scale problems. The computational bound of each iteration in StOMP is $\mathcal{O}(MN)$.

4.5.1.4 Regularized Orthogonal Matching Pursuit

Regularized orthogonal matching pursuit (ROMP) is a modified extension of OMP [63, 64]. ROMP provides a strong theoretical guarantee for sparse signal recover if the sensing matrix $\boldsymbol{A}$ satisfies RIP. Different from OMP, ROMP requires prior knowledge about the sparsity S of $\boldsymbol{x}$. To deal with this problem, one can conduct empirical studies using different sparsity levels and selects the one that minimizes $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2$.

Similar to StOMP, the correlations between the columns of $\boldsymbol{A}$ and the redidual are calculated as in (4.20). At each iteration, ROMP selects S columns that have the largest correlation with the redidual, or all the columns with non-zero correlation, whichever set is smaller. Let I_s denote the selected columns, and I_r is a subset of I_s that satisfies

$$c_i \leq 2c_j, \forall\, i, j \in I_0, \tag{4.25}$$

where c_i and c_j are as defined in (4.20). Then, choose I_r such that

$$I_r = \operatorname{Arg}\max_{I_r} \sum_{I_r} c_i^2, \quad i \in I_r. \tag{4.26}$$

The set of columns selected to represent the signal is updated as follows:

$$T^{(j)} = T^{(j-1)} \cup I_r. \tag{4.27}$$

Once the columns are updated, new estimator $\boldsymbol{x}_T^{(j)}$ is calculated as in (4.23). ROMP stops when the iteration number reaches the sparsity of the signal S, or the number of entries in $T^{(j)}$ is larger than $2S$.

When the sensing matrix $\boldsymbol{A}$ has the Restricted Isometry constant satisfying $\delta_{4S} < 0.01/\sqrt{\log S}$, the ROMP is guaranteed to recover the exact sparse signal if no noise is present [63, 64]. In the noisy setting, Then ROMP produces an approximation $\hat{\boldsymbol{x}}$ to S-sparse signal $\boldsymbol{x}$ that satisfies:

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq 104 \log S \|\boldsymbol{\epsilon}\|_2.$$

Each iteration of ROMP requires calculation of the correlation between the residual and the columns in $\boldsymbol{A}$, which needs running time bound of $\mathcal{O}(MN)$. The selection of the index set I_s can be done by sorting the value of c_i in a nonincreasing order, and selecting S biggest entries. Many sorting algorithms such as Mergesort or Heapsort provide running times of $\mathcal{O}(N \log N)$. Selecting a subset I_r from I_s as in (4.26) requires running time $\mathcal{O}(\log S)$. In the next step, solving $\boldsymbol{x}_T$ in (4.23) requires $\mathcal{O}(S^2 M)$ operations in general, and $\mathcal{O}(SM)$ operations when $\boldsymbol{A}$ is incoherent or satisfies RIP. ROMP terminates in at most $2S$ iterations. Therefore, in case $M \geq \log N$ and $N \geq S^2$, the total running time of ROMP is $\mathcal{O}(SMN)$. This is the same bound as that of OMP.

4.5.1.5 Acrostic Compressive Sampling Matching Pursuit

Inspired from ROMP, *acrostic compressive sampling matching pursuit* (CoSaMP) is another iterative greedy algorithm which seeks a sparse solution from an underdetermined linear system [65]. CoSaMP is based on OMP. It accelerates OMP, and therefore, it is computationally more efficient [65].

At each iteration, CoSaMP selects $2S$ columns that have the largest correlation with the redidual, and the index set $T^{(j)}$ is updated as follows,

$$T^{(j)} = \text{support of } \boldsymbol{x}^{(j-1)} \cup I_s. \tag{4.28}$$

$\boldsymbol{x}_T$ is estimated as

$$\boldsymbol{x}_T^{(j)} = (\boldsymbol{A}_T^T \boldsymbol{A}_T)^{-1} \boldsymbol{A}_T^T \boldsymbol{y}. \tag{4.29}$$

Then, only the S largest coefficients in $\boldsymbol{x}_T^{(j)}$ are kept and the rest of the coefficients are pruned to zero. CoSaMP stops after a fixed number of iterations, or when the residual or the correlation between the columns $\boldsymbol{A}$ and the residual is small enough.

If the sensing matrix $\boldsymbol{A}$ has restricted isometry constant $\delta_{2S} \leq 1$, for a given precision parameter η, the algorithm CoSaMP produces an S-sparse approximation $\hat{\boldsymbol{x}}$ that satisfies [65]

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq C \cdot \max\{\eta, \frac{1}{\sqrt{S}} \|\boldsymbol{x} - \boldsymbol{x}_{S/2}\|_1 + \|\boldsymbol{\epsilon}\|_2\}.$$

In the absence of noise, CoSaMP can recover $\boldsymbol{x}$ with an arbitrarily high precision.

Similar as OMP, the major obstacle to fast implementation of CoSaMP comes from calculation of the correlation between the residual and the columns in $\boldsymbol{A}$ and the least square estimation in (4.29). The computational bound of each iteration in CoSaMP is $\mathcal{O}(MN)$.

4.5.2 Convex Relaxation

The optimization problems in (4.5) and (4.7) are NP hard. Therefore, the l_0-norm constraint is often relaxed to some convex sparsity constraint $f(\boldsymbol{x})$.

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{\Gamma} = \boldsymbol{A}\boldsymbol{x}. \tag{4.30}$$

The most popular choice of $f(\boldsymbol{x})$ is the following l_1-norm constraint,

$$\min_{\boldsymbol{x}} \; \|\boldsymbol{x}\|_1 \quad \text{such that} \quad \boldsymbol{\Gamma} = \boldsymbol{A}\boldsymbol{x}. \tag{4.31}$$

The optimization problem in (4.31) is also known as the basis pursuit (BP) [67]. BP poses a convex optimization problem, and can be solved by a number of linear programming techniques [67]. In the noisy setting, BP is reformulated as a denoising problem as follows:

$$\min_{\boldsymbol{x}} \; \|\boldsymbol{x}\|_1 \quad \text{such that} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \varepsilon. \tag{4.32}$$

The solution to (4.32) differs from the true sparsest solution by at most a constant multiple of the noise level [55]. The solution to BP is based on decomposing a signal into an optimal superposition of elements which have the smallest l_1-norm of coefficients.

BP is closely related to *least absolute shrinkage and selection operator* (LASSO) approach, which has the following form,

$$\min \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \quad \text{such that} \quad \|\boldsymbol{x}\|_1 < t, \tag{4.33}$$

where t is some predetermined threshold value. Another variation of BP and LASSO is to solve an optimization problem by trading off a quadratic error of the measurements with an l_1-norm of the signal,

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1, \tag{4.34}$$

where λ is the regularization parameter.

The convex relaxation are closely related to linear programming, and can be solved by a number of algorithms. We briefly review some these algorithms in the following subsections.

4.5.2.1 BP-simplex

The simple algorithm to solve BP is the BP-simplex [67]. The sensing matrix $\boldsymbol{A}$ is assumed to be an $M \times N$ matrix with $M \ll N$. The algorithm first chooses M linearly independent columns from $\boldsymbol{A}$. Let T denote the index set of the M selected columns and $\boldsymbol{A}_T$ be the submatrix which consists of the M selected columns. The initial solution is then

$$\boldsymbol{x}_T = \boldsymbol{A}_T^{-1}\boldsymbol{y}. \tag{4.35}$$

Next, the algorithm iteratively improves the current basis by swapping one selected column of $\boldsymbol{A}_T$ with the one outside $\boldsymbol{A}_T$ so that $\|\boldsymbol{x}\|_1$ is reduced at each step. It was shown that there is always a swap that improves or maintains the objective value except at the optimal solution [67].

Each iteration of BP-simplex requires computing the inverse problem (4.35). The inversion of the $M \times M$ matrix $\boldsymbol{A}$ costs running time of $\mathcal{O}(M^3)$. If the columns of $\boldsymbol{A}$ are incoherent, some iterative methods such as CG can reduce this bound to $\mathcal{O}(M^2)$. The number of iterations in BP-simplex is limited by the number of columns in $\boldsymbol{A}$.

4.5.2.2 BP-interior

The formula in (4.34) is an l_1-norm regularized least squares problem which is convex quadratic, and can be solved by standard *interior point methods* [67,129,130].

The implementation of *interior point methods* requires calculation of a lower bound of the objective function (4.34). To do that, first a dual feasible point $\boldsymbol{\nu}$ is introduced,

$$\boldsymbol{\nu} = 2s(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}), \quad s = \min_i \frac{\lambda}{|2(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})_i|}. \tag{4.36}$$

Let $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}$. The Lagrangian of (4.34) associated with the dual variable $\boldsymbol{\nu}$ is

$$L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}) = \boldsymbol{z}^T\boldsymbol{z} + \lambda\|\boldsymbol{x}\|_1 + \boldsymbol{\nu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y} - \boldsymbol{z}). \tag{4.37}$$

Thus, the inferior of (4.37) is given by

$$\inf_{\boldsymbol{x},\boldsymbol{z}} L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{\nu} - \boldsymbol{\nu}^T\boldsymbol{y}. \tag{4.38}$$

$\inf_{\boldsymbol{x},\boldsymbol{z}} L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu})$ gives a lower bound of the objective function (4.37), and the difference between the primary objective function (4.34) and the lower bound is called the duality gap, denoted by η,

$$\eta = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 - \inf_{\boldsymbol{x},\boldsymbol{z}} L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}). \tag{4.39}$$

The optimization problem in (4.34) can be equivalently written as a quadratic optimization problem with inequality constraint,

$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\sum_{i=1}^{N} u_i, \quad \text{such that} \quad -u_i \leq x_i \leq u_i,\ i = 1, ...N, \tag{4.40}$$

where $u_i \geq 0$ $(i = 1..., N)$ are constraints for x_i. A logarithmic barrier for the bound constraints is defined as follows:

$$\Xi(\boldsymbol{x}, \boldsymbol{u}) = -\sum_{i=1}^{N}\log(u_i + x_i) - \sum_{i=1}^{N}\log(u_i - x_i). \tag{4.41}$$

$\Xi(\boldsymbol{x}, \boldsymbol{u})$ approaches $+\infty$ when $|x_i|$ approaches the bound u_i. The solution of (4.40) can be obtained by minimizing the following objective function

$$\Phi(\boldsymbol{x}, \boldsymbol{u})_p = p(\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\sum_{i=1}^{N} u_i) + \Xi(\boldsymbol{x}, \boldsymbol{u}). \tag{4.42}$$

where the parameter $p > 0$ is a constant parameter. As p increases, the minimizer of (4.42) approaches the solution of (4.40). (4.42) is a convex and differentialable function with unique minimizer at $(\boldsymbol{x}_p, \boldsymbol{u}_p)$, which can be solved by the gradient approach, such as *Newton's method.* The algorithm seeks a sequence of minimizers $(\boldsymbol{x}_p, \boldsymbol{u}_p)$ of increasing p value, and terminates when the duality gap in (4.39) is small enough.

For a large scale problem, minimizing (4.42) by Newton's method is computationally expensive. For large scale problems, *interior point methods* use *conjugate gradient* (CG) [130], or the *preconditioned conjugate gradient* (PCG) methods [129] for efficient implementation.

The computational complexity of BP-interior depends on the duality gap η. When η is small, the minimizer of (4.42) is close to the solution of (4.40). Therefore, the value of p in (4.42) is large, and the number of log-barrier iterations goes up. Each log-barrier iteration seeks the minimizer of (4.42) by iterative gradient approach, such as CG. Each iteration of CG has computational complexity of $\mathcal{O}(MN)$. In [129], a preconditioner was used to accelerate the convergence rate of CG and shows a great reduction in running time for large scale problems. The empirical complexity BP-interior using PCG iteration is $\mathcal{O}(N^{1.2})$ [129].

4.5.2.3 Iterative Shrinkage/Thresholding

The *iterative shrinkage/thresholding algorithm* (IST) is an iterative algorithm that amounts to a Landweber iteration with shrinkage or thresholding at each iteration [68–70].

The IST algorithm first picks a constant $C > \|\boldsymbol{A}\|_2$, where $\|\cdot\|_2$ denotes the spectral norm [1], and then defines a function which depends on an auxiliary vector $\boldsymbol{u}$ as follows:

$$\Xi(\boldsymbol{x}, \boldsymbol{u}) = C\|\boldsymbol{x} - \boldsymbol{u}\|_2^2 - \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{u})\|_2^2. \tag{4.43}$$

Since $C > \|\boldsymbol{A}^T\boldsymbol{A}\|_2$, $\Xi(\boldsymbol{x}, \boldsymbol{u})$ is strictly convex in $\boldsymbol{x}$ with any choice of $\boldsymbol{u}$. Without loss of generality, assuem $\|\boldsymbol{A}\|_2 < 1$ and set $C = 1$, since $\boldsymbol{A}$ can always be normalized.

---

[1] The spectral norm of a matrix $\boldsymbol{A}$ is the largest singular value of $\boldsymbol{A}$, or the square root of the largest eigenvalue of the positive-semidefinite matrix $\boldsymbol{A}^T\boldsymbol{A}$

By adding (4.43) to (4.34), a surrogate function is defined as follows:

$$\begin{aligned}\Phi(\boldsymbol{x},\boldsymbol{u}) &= \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{x}\|_1+C\|\boldsymbol{x}-\boldsymbol{u}\|_2^2-\|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{u})\|_2^2\\ &= \|\boldsymbol{x}\|_2^2-2\boldsymbol{x}^T(\boldsymbol{u}+\boldsymbol{A}^T\boldsymbol{y}-\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u})+\lambda\|\boldsymbol{x}\|_1+\|\boldsymbol{y}\|_2^2+\|\boldsymbol{u}\|_2^2\\ &\quad -\|\boldsymbol{A}\boldsymbol{u}\|_2^2 \qquad (4.44)\end{aligned}$$

Since $\Xi(\boldsymbol{x},\boldsymbol{u})$ is strictly convex in $\boldsymbol{x}$, $\Phi(\boldsymbol{x},\boldsymbol{u})$ is also strictly convex in $\boldsymbol{x}$ with any $\boldsymbol{u}$. Compared with the original objective function in (4.34), (4.44) removes the quadratic term $\|\boldsymbol{A}\boldsymbol{x}\|_2^2$ which leads to simplicity in implementation.

$\Phi(\boldsymbol{x},\boldsymbol{u})$ is differentialable at x_i if $x_i \neq 0$, the gradient of $\Phi(\boldsymbol{x},\boldsymbol{u})$ with respect to x_i is given as

$$\nabla_{x_i}\Phi(\boldsymbol{x},\boldsymbol{u})x_i=(\boldsymbol{u}+\boldsymbol{A}^T\boldsymbol{y}-\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u})_i+\lambda/2\text{sign}(x_i),\quad \forall\, 1\le i\le N,\ x_i\neq 0. \quad (4.45)$$

Thus, since (4.44) is convex, the solution to (4.44) is given when the gradient of $\Phi(\boldsymbol{x},\boldsymbol{u})$ with respect to $\boldsymbol{x}$ equals zero, and the $\boldsymbol{x}$ that minimize (4.44) is given by

$$x_i=\begin{cases}(\boldsymbol{u}+\boldsymbol{A}^T\boldsymbol{y}-\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u})_i-\lambda/2 & (\boldsymbol{u}+\boldsymbol{A}^T\boldsymbol{y}-\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u})_i>\lambda/2\\ (\boldsymbol{u}+\boldsymbol{A}^T\boldsymbol{y}-\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u})_i+\lambda/2 & (\boldsymbol{u}+\boldsymbol{A}^T\boldsymbol{y}-\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u})_i<-\lambda/2\\ 0 & \text{otherwise.}\end{cases} \quad (4.46)$$

IST solves (4.44) iteratively, the auxiliary variable at the jth iteration $\boldsymbol{u}^{(j)}$ is chosen as the solution $\boldsymbol{x}^{(j-1)}$ from the last iteration,

$$\boldsymbol{x}^{(j)}=\text{Arg}\min_{\boldsymbol{x}}\Phi(\boldsymbol{x},\boldsymbol{x}^{(j-1)}). \quad (4.47)$$

Starting from an arbitrary $\boldsymbol{u}^{(0)}$, the sequence $\boldsymbol{x}^{(j)}, j=1,2,...$ converges to the unique minimizer of (4.34) [69].

From (4.46), the main computation effort in each iteration of IST comes from the the matrix-matrix multiplication $\boldsymbol{A}^T\boldsymbol{A}$, and matrix-vector multiplication $\boldsymbol{A}^T\boldsymbol{y}$, which has running time of $\mathcal{O}(MN^2)$ and $\mathcal{O}(MN)$, respectively. The thresholding step uses a running time of $\mathcal{O}(N)$. Therefore, the overall computational bound of each iteration in IST is $\mathcal{O}(MN^2)$.

4.5.2.4 Gradient Projection for Sparse Reconstruction

Gradient projection for sparse reconstruction (GPSR) seeks a solution of (4.34) iteratively. It projects the negative-gradient direction of the objective function onto the solution from last iteration [71].

The GPSR approach splits the variable $\boldsymbol{x}$ into its positive and negative parts,

$$\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v}, \quad \boldsymbol{u} \geq 0,\ \boldsymbol{v} \geq 0. \tag{4.48}$$

These relationships are satisfied by $u_i = (x_i)_+$ and $v_i = (-x_i)_+$ for all $i = 1, 2, ..., N$, where $(\cdot)_+$ denotes the positive-part operator defined as $(x_i)_+ = \max\{0, x_i\}$. Thus, $\|\boldsymbol{x}\|_1$ can be written in the following equation,

$$\|\boldsymbol{x}\|_1 = \mathbf{1}^T\boldsymbol{u} + \mathbf{1}^T\boldsymbol{v}, \tag{4.49}$$

where $\mathbf{1}$ is the vector consisting of N ones. The objective function in (4.34) can be rewritten as the following bound-constrained quadratic program (BCQP):

$$\min_{\boldsymbol{u},\boldsymbol{v}} \quad \boldsymbol{F}(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{v})\|_2^2 + \lambda\mathbf{1}^T\boldsymbol{u} + \lambda\mathbf{1}^T\boldsymbol{v}, \quad \text{such that} \quad \boldsymbol{u} \geq 0, \boldsymbol{v} \geq 0. \tag{4.50}$$

where $\boldsymbol{z} = [\boldsymbol{u}^T, \boldsymbol{v}^T]^T$.

[71] proposes two ways to update $\boldsymbol{z}$ iteratively and solve (4.50): (1) GPSR-Basic algorithm, and (2) GPSR-BB algorithm. The first algorithm searches from each iterate $\boldsymbol{z}^{(j)}$ along the negative gradient

$$\boldsymbol{z}^{(j+1)} = (\boldsymbol{z}^{(j)} - \alpha^{(j)}\nabla F(\boldsymbol{z}^{(j)}))_+. \tag{4.51}$$

The $(\cdot)_+$ keeps the solution nonnegative. The second algorithm replaces $\nabla F(\boldsymbol{z}^{(j)})$ in (4.51) by $\boldsymbol{H}^{-1}(\boldsymbol{z}^{(j)})\nabla F(\boldsymbol{z}^{(j)}))$,

$$\boldsymbol{z}^{(j+1)} = (\boldsymbol{z}^{(j)} - \alpha^{(j)}\boldsymbol{H}^{-1}(\boldsymbol{z}^{(j)})\nabla F(\boldsymbol{z}^{(j)})))_+, \tag{4.52}$$

where $\boldsymbol{H}(\boldsymbol{z}^{(j)})$ is an approximation to the Hessian of $F(\boldsymbol{z}^{(j)})$. $\boldsymbol{H}(\boldsymbol{z}^{(j)})$ can be calculated efficiently by the Barzilai and Borwein (BB) approach [131]. The solution

$\boldsymbol{x}$ is derived directly from $\boldsymbol{z}$.

It is not possible to accurately predict the number of GPSR-Basic and GPSR-BB iterations required to find an approximate solution. In each iteration, the main computational cost requires calculation of $\nabla F(\boldsymbol{z}^{(j)})$ as in (4.51) for GPSR-Basic, or $\boldsymbol{H}^{-1}(\boldsymbol{z}^{(j)})\nabla F(\boldsymbol{z}^{(j)})$ as in (4.52) for GPSR-BB. A number of matrix-matrix multiplication and matrix-vector multiplication and are needed. The computational bound in each iteration of GPSR is $\mathcal{O}(MN^2)$.

4.5.2.5 Focal Underdetermined System Solver

Focal underdetermined system solver (FOCUSS) is an iterative reweighted minimum norm algorithm [72, 73] which is able to solve the l_1-norm constraint optimization problem. It first derives a low resolution initial estimate of the signal, then refines the initial estimator iteratively by obtaining a reweighted minimum norm solution.

For an underdetermined system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, the minimum norm solution, i.e., the solution with minimum $\|\boldsymbol{x}\|_2$, is given by

$$\boldsymbol{x} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{y} = \boldsymbol{A}^{\dagger}\boldsymbol{y}, \tag{4.53}$$

where $\boldsymbol{A}^{\dagger} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$ is the Moore-Penrose pseudoinverse.

In FOCUSS, the objective is to find a solution with minimum weighted norm

$$\boldsymbol{x} = \operatorname{Arg}\min_{\boldsymbol{x}} \|\boldsymbol{W}^{-1}\boldsymbol{x}\|_2, \quad \text{such that} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{4.54}$$

where $\boldsymbol{W}$ is the weight matrix. When $\boldsymbol{W}$ is diagonal with diagonal entries equals to $\boldsymbol{x}$, (4.54) is equivalent to an l_1-norm constraint optimization problem. The solution to (4.54) is given by

$$\boldsymbol{x} = \boldsymbol{W}(\boldsymbol{A}\boldsymbol{W})^{\dagger}\boldsymbol{y}. \tag{4.55}$$

In basic FOCUSS, $\boldsymbol{W}^{(j)}$ is a diagonal matrix with diagonal entries equal to the solution from the last iteration, i.e., $\boldsymbol{x}^{(j-1)}$. Thus, the algorithm is able to reinforce large entries in $\boldsymbol{x}^{(j-1)}$ in the following iteration while suppressing the small entries. The zero entries in $\boldsymbol{W}^{(j)}$ can be eliminated from the product $\boldsymbol{A}\boldsymbol{W}$. However, the matrix $(\boldsymbol{AW})^\dagger$ might be ill-conditioned, Tikhonov regularization or singular value decomposition might be used to calculate (4.55). In FOCUSS, an initial estimate which is close to the final solution improves the convergence rate and the final estimate [72, 73].

The computational effort of FOCUSS involves solving a sequence of quadratic least square problems. In each iteration, the weight matrix $\boldsymbol{W}$ can be easily derived from the solution of the last iteration, and the computational time is $\mathcal{O}(N)$. The heaviest computational burden in each iteration is calculating the pseudoinverse in (4.55), which needs a computational time of $\mathcal{O}(M^2N)$. Therefore, the total computational bound in each iteration of FOCUSS is $\mathcal{O}(M^2N)$. The number of iterations in FOCUSS depends on the initial estimate. If the initial estimate is close to the final solution, only a small number of iterations are needed for convergence [73].

4.5.3 Non-convex Relaxation

Non-convex sparsity constraints are used to relax with the l_p-norm, $0 \leq p < 1$, constraint,

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_p \quad \text{such that} \quad \|\boldsymbol{y} - \boldsymbol{Ax}\|_2 \leq \varepsilon. \tag{4.56}$$

In CS theory, it was shown that l_p-norm, $0 \leq p < 1$, constraint optimization leads to the exact solution. Some recent research shows that the l_p-norm constraint optimization is able to reconstruct signal with fewer measurements compared with l_1-norm constraint optimization (4.32) [74]. The optimization problem in (4.56) can be solved by FOCUSS in a similar fashion as described in Section 1.5.2.4, with the diagonal entries of $\boldsymbol{W}^{(j)}$

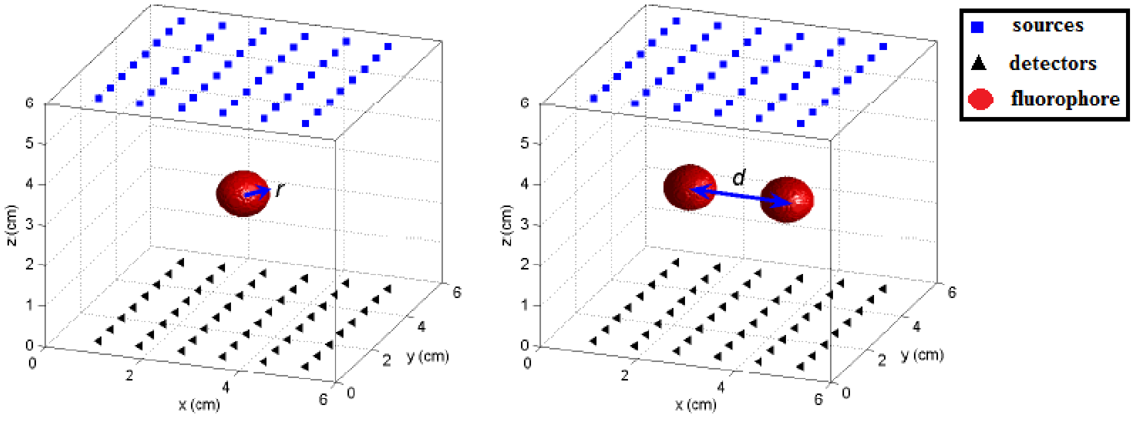
$$w_i^{(j)} = |x_i^{(j-1)}|^{2-p}, \quad \text{for } 1 \leq i \leq N, \tag{4.57}$$

where $\boldsymbol{x}^{(j-1)}$ is the solution at the $j-1$th iteration. Note that with this choice of $\boldsymbol{W}$, the constraint term $\|\boldsymbol{W}^{-1}\boldsymbol{x}\|_2$ is $\|\boldsymbol{x}\|_p$.

4.6 Numerical Simulations

This section uses numerical simulation to test the performances of various sparse promoting reconstruction algorithms in the FDOT inverse problem. For comparison purpose, we also give the reconstruction results by commonly used Tikhonov regularization. Both visual results and quantitative measurements are given to evaluate the performance different reconstruction algorithms.

4.6.1 Simulation Setup



(a) (b)

Figure 4.1: Simulation setup.

The simulated imaging domain is a 6 cm × 6 cm × 6 cm cubic region, with 48 sources and 48 detectors uniformly distributed on the top and bottom surfaces, respectively, as shown in Figure 4.1. The fluorophore is concentrated in the red spherical region. We set the background absorption coefficient to $\mu_{am} = 0.05$ cm^{-1}, and the diffusion coefficient to $D(\boldsymbol{r}) = 0.04$ cm^{-1} at both the excitation and emission wavelengths. The circular heterogeneity with radius r simulates the fluorophore concentration with fluorophore absorption coefficient $\mu_{axf} = 0.005$ cm^{-1}. In Figure 4.1(a), a single spherical heterogeneity is located at the center of the imaging domain. In Figure 4.1(b), there are two separated spherical heterogeneities of the same sizes, and the distance between their centers is $d$cm.

We discretized the imaging domain into $20 \times 20 \times 20$ voxels. Thus, the forward sensing matrix is 2304 by 8000. For the phantom with single heterogeneity, we simulated different data sets corresponding to 3 different values of the radius r : 0.5, 0.75, and 1.0 cm. For the phantom with double heterogeneities, we simulated 3 different distances d between two heterogeneities: 1.0, 1.5, and 2.0 cm.

To simulate the measurement noise, we assumed that, when a sufficiently large number of photons are detected, the noise at each detector can be approximated by a Gaussian random variable with its variance proportional is to the magnitude of the detector reading. The SNR of the measurements is defined as

$$SNR = 20 \log_{10} \frac{\|\mathbf{\Gamma}\|_2}{\|\boldsymbol{\epsilon}\|_2}. \tag{4.58}$$

We simulated three sets of noise contaminated measurements with approximately 10%, 3% and 1% noise, corresponding to the SNR value of 20 dB, 30 dB and 40 dB, respectively.

4.6.2 Imaging Reconstruction - Visual Results

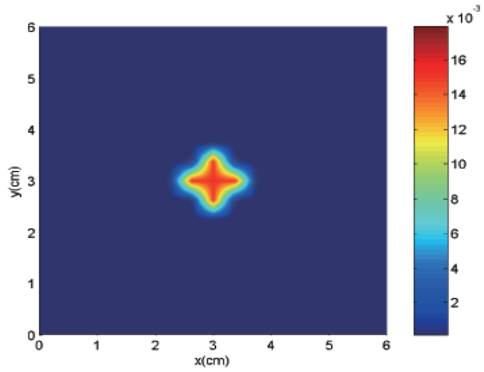


Figure 4.2: The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$ cm).

In this simulation, we first reconstructed the fluorophore concentration of the phantom with single heterogeneity, as shown in Figure 4.1(a). The fluorophore heterogeneity is set at 3 different sparsity levels corresponding to 3 different radii. Next, we reconstructed the fluorophore concentration of the phantom with two heterogeneities, and simulated 3 different distance between the two heterogeneities. In each case, we used different sparsity promoting reconstruction methods including both greedy algorithms and relaxation techniques. The reconstruction results of Tikhonov regularization is also presented for comparative purpose.

Figure 4.2 shows the cross sections of the simulated phantom with a single heterogeneity of radius 0.5cm. The reconstruction results using different algorithms are shown in Figure 4.3. Figure 4.3(b) shows the reconstruction result of the commonly used Tikhonov regularization method. The result has a severe oversmoothing effect. Figure 4.3(a)-4.3(f) are the reconstruction results by different greedy algorithms. In all of these figures, the reconstructed fluorophore maps appear much smaller than the original simulated phantom, and large spikes of noise are presented. In general, the reconstruction results of greedy algorithms do not look as good as those of relaxation techniques as in Figure 4.3(g)-4.3(l). Because accurate reconstruction from greedy algorithms typically requires the columns of the sensing matrix highly incoherent to each other. However, in FDOT imaging problem, the diffusive nature of NIR light makes the columns of the forward matrix less incoherent. Comparatively, although relaxation algorithms also benefit from an incoherent sensing matrix, they work well with less incoherent matrix as well. Relaxation algorithms have more accurately reconstructed the size and the location of the heterogeneity, and less background noise is present.

Figure 4.4 shows the cross sections of the simulated phantom with two heterogeneities in the imaging domain. The diameter of each heterogeneity is 1.0cm and the distance between the centers of the two heterogeneities is 1.5cm. Figure 4.5 demonstrates how different reconstruction algorithms can differentiate multiple heterogeneities in the imaging domain. The reconstruction result of the commonly used Tikhonov regularization is shown in Figure 4.5(a), where severer oversmoothing makes the two heterogeneities hard to differentiate. The reconstruction results from sparse signal recovery algorithms show two separated heterogeneities in the reconstructed images. The reconstruction results of greedy algorithms have large background noise in the images, as shown in Figure 4.5(b)-4.5(f). The reconstruction results of relaxation algorithms, as shown in Figure 4.5(g)-4.5(l), seem to have better agreement with the original phantom.

Table 4.1: NMSE of the reconstructed images (10^{-4}) using different sparse signal recovery techniques.

Algorithm	$r = 0.5$ cm			$r = 0.75$ cm			$r = 1.0$ cm		
SNR (dB)	40	30	20	40	30	20	40	30	20
Tikhonov	13.3	20.7	31.7	19.0	29.1	48.4	27.8	79.1	129
OMP	5.02	5.92	19.0	8.03	17.8	29.6	19.7	50.3	70.1
GP	6.08	8.03	15.9	5.21	15.3	33.1	13.6	36.7	68.9
StOMP	2.05	6.47	12.1	3.87	14.5	28.0	9.44	31.8	63.7
ROMP	1.66	5.85	13.2	3.68	13.0	31.6	8.01	29.0	64.2
CoSaMP	2.11	6.35	13.0	5.97	14.5	28.9	11.8	34.7	67.8
BP-simplex	2.31	4.21	10.8	6.22	8.08	20.3	7.99	30.0	49.2
BP-interior	1.56	3.03	7.05	2.44	7.13	15.3	5.03	20.8	39.1
IST	1.20	3.55	7.49	3.51	6.97	19.2	6.38	23.3	49.9
GPSR	1.29	3.28	7.07	3.41	7.13	17.1	5.51	22.2	46.6
Focuss(l_1)	1.81	4.31	9.89	3.01	8.37	29.8	6.00	28.9	63.8
Focuss(l_p)	1.95	3.76	11.1	4.08	7.38	22.3	4.98	27.4	59.0

4.6.3 Imaging Reconstruction - Quantitative Measurements

We measured the accuracy of the reconstructed images using the normalized mean square error (NMSE) given by

$$\text{NMSE} = \frac{\|\boldsymbol{x}_{\text{true}} - \boldsymbol{x}_{\text{recon}}\|_2}{\# \text{ of voxels}}. \tag{4.59}$$

We averaged the NMSE of the reconstructed images over 50 realizations of noise. The results were tabulated in Table 4.1. We observed that all sparse promoting algorithms has a lower NMSE compared to the Tikhonov regularization. The NMSE increases when the radius of the heterogeneity increases. For different sparse signal recovery algorithms, the NMSE from relaxation algorithms are smaller than that of greedy algorithms in general. However, greedy algorithms have the advantage of straight-forwardness, which are typically faster and easier in implementation.

To quantify the resolvability of two closely spaced heterogeneities in the reconstructed images, we referred to the *peak-to-valley-ratio* (PVR). PVR is defined as the ratio between the average reconstructed absorption coefficient of the fluorophore at the two peak values corresponding to the true fluorophore locations and that at the lowest value between the two peaks. We summarized the PVR of the reconstructed images in Table 4.2. PVR results show that the reconstructed images

Table 4.2: PVR of the reconstructed images using different sparse signal recovery techniques.

Algorithm	$d = 1.0$ cm			$d = 1.5$ cm			$d = 2.0$ cm		
SNR (dB)	40	30	20	40	30	20	40	30	20
Tikhonov	1.87	1.22	0.97	2.53	2.37	2.30	2.73	2.32	2.52
OMP	2.00	1.00	0.84	3.06	2.26	1.93	6.59	5.78	4.90
GP	1.64	0.73	0.67	2.54	1.48	1.33	6.94	4.35	3.87
StOMP	4.32	3.33	3.13	8.38	6.50	5.06	20.0	16.0	11.4
ROMP	2.53	1.56	1.31	12.4	11.8	8.57	25.7	32.2	21.9
CoSaMP	3.66	2.93	2.73	47.8	35.6	33.1	98.2	69.0	31.0
BP-simplex	2.96	2.54	2.33	13.7	10.8	8.97	39.9	31.5	21.3
BP-interior	2.11	1.48	1.36	28.8	22.1	18.7	55.1	47.1	40.0
IST	4.43	3.90	3.65	23.8	19.1	15.5	103	65.5	50.8
GPSR	5.49	4.53	4.38	27.7	26.1	20.3	41.4	40.1	38.8
Focuss(l_1)	8.51	3.36	2.15	102	50.8	18.9	123	89.3	30.9
Focuss(l_p)	6.69	3.09	2.96	113	46.9	31.1	175	97.8	42.1

by the relaxation algorithms generally have higher resolvability compared to those by the greedy-type algorithms. The sparse signal recovery algorithms provide better PVR compared to the traditional Tikhonov regularization.

4.7 Silicon Phantom Experiment

Figure 4.6(a) shows an illustration of the cylindrical phantom used in the experiment. The phantom was made of silicone rubber with diameter of about 2 cm, and length of 4 cm. The phantom had homogeneous absorption coefficient $\mu_a = 0.2$ cm^{-1} and scattering coefficient $\mu_s' = 12$ cm^{-1} ($D = 1/3(\mu_a + \mu_s')$) at both the excitation and emission wavelengths (743 nm and 767 nm). The silicon phantom contained a hollow cylindrical tube in the middle with approximately 3 mm in diameter. which was filled with intralipid and ink to mimic the same optical properties as the background. The intralipid and ink contained 1 micromolar of Cy7 as the fluorophore. The cross section of the fluorophore yield at $z = 1$ is shown in Figure 4.6(b).

The FDOT measurements were collected using the FDOT imaging system reported in [132], as illustrated in Figure 4.7. Specifically, focused collimated laser beams were used as point light sources to excite the fluorophore. We had 60 point sources in total. The fluorescence measurements were collected by an electrically cooled CCD camera. The reading of the detector was recorded as the mean value of a subregion with 5×5 pixels around each detector location. We selected 60 detector locations. We discretized the imaging domain into $20 \times 20 \times 20$ voxels. Thus, the forward sensing matrix is of dimension 3600 by 8000.

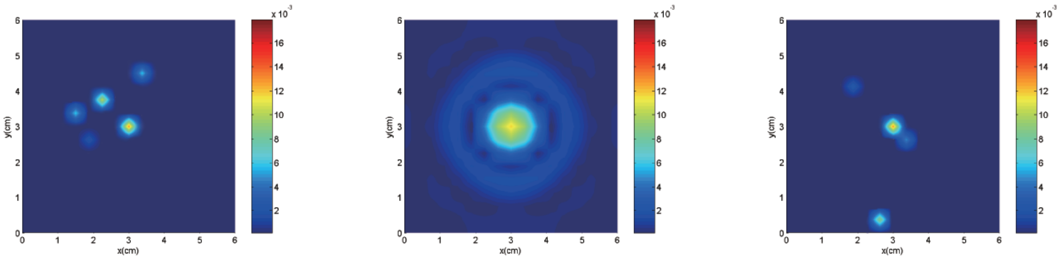
Figure 4.8 shows the cross sections of the real phantom and the reconstruction results using different reconstruction techniques. Compared with Figure 4.6(b), the reconstruction results of the Tikhonov regularization in Figure 4.8(a) shows severer oversmoothing, which is similar as the visual results in the numerical simulation. The reconstruction from greedy algorithms are in Figure 4.8(b)-4.8(f). The images are not oversmoothing, but look discrete and large noise is present in the background. The reconstruction results of relaxation algorithms are in Figure 4.8(g)-4.8(l). The images have small oversmoothing effect at the boundary of the fluorophore. The reconstruction results of the relaxation algorithms look better than that of the greedy algorithms according to the real phantom in Figure 4.6(b).

4.8 Discussion

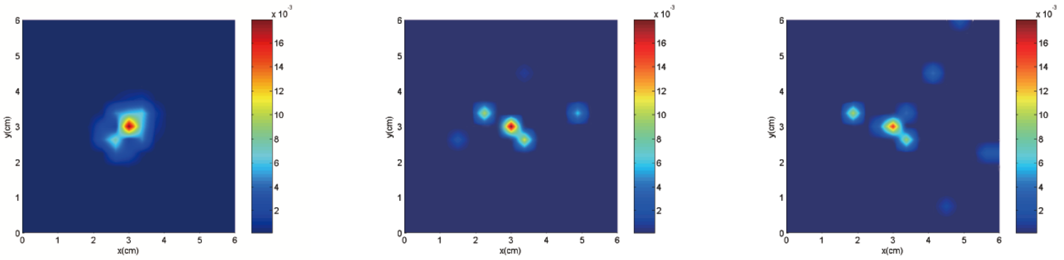
In this chapter, we reviewed a number of sparse signal recovery techniques available in the literature. In Section 3.3 of Chapter 3, we introduced a combined l_2-l_1-norm regularization technique for FDOT image reconstruction. We note that the combined l_2-l_1-norm regularization can be considered as a sparse signal recovery technique combined with specific *a priori* information. The recovery of the foreground region of the imaging domain was regularized by the l_1-norm constraint, and the l_2-norm constraint smoothes out the background noise and suppresses large spikes of artifacts due to the ill-posedness of the FDOT forward sensing matrix. The performances of FDOT reconstruction using l_1-norm regularization and the combined l_2-l_1-norm regularization were presented in numerical simulation and using real data obtained from a silicon phantom experiment in Chapter 3. The results indicate that the application of prior information combined with sparsity promoting regularization improves the quantitative accuracy of FDOT image reconstruction.

4.9 Conclusion

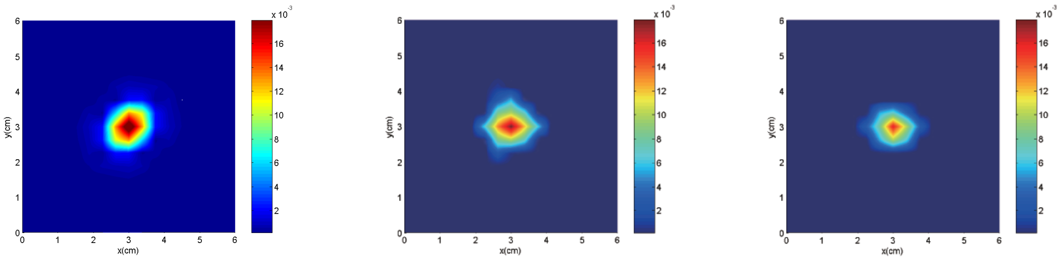
This chapter applied the sparse signal techniques from CS literature to address the FDOT inverse problem. The basic theory of CS and a number of sparse promoting algorithms were reviewed. We reconstructed the fluorophore yield in FDOT inverse problem by various sparse promoting techniques in CS literature. We used both numerical simulation and real phantom data to test the performances of different algorithms, and compared them with the traditional Tikhonov regularization. We showed that in general, the sparse promoting algorithms provide better reconstruction results than the Tikhonov regularization. The relaxation algorithms usually have better performance compared to the greedy algorithms. Because the accurate reconstruction of greedy algorithms typically requires the columns of the sensing matrix highly incoherent, however, the diffusive nature of the NIR light in the imaging domain makes the columns in FDOT forward matrix less incoherent. Compared to the greedy algorithms, the relaxation algorithms are less dependent on the incoherency of the sensing matrix, thus can work better with FDOT forward matrix.



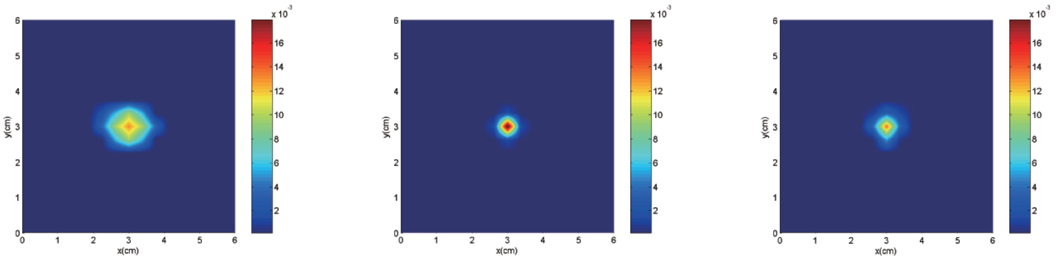
(a) OMP (b) Tikhonov (c) GP



(d) StOMP (e) ROMP (f) CoSaMP



(g) BP-simplex (h) BP-interior (i) IST



(j) GPSR (k) FOCUSS (l_1) (l) FOCUSS (l_p)

Figure 4.3: The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using different algorithms, $r = 0.5$ cm, 1% noise.

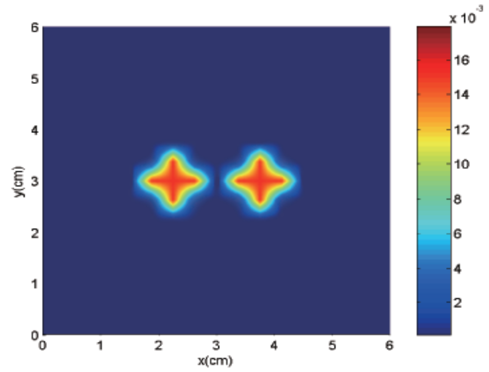
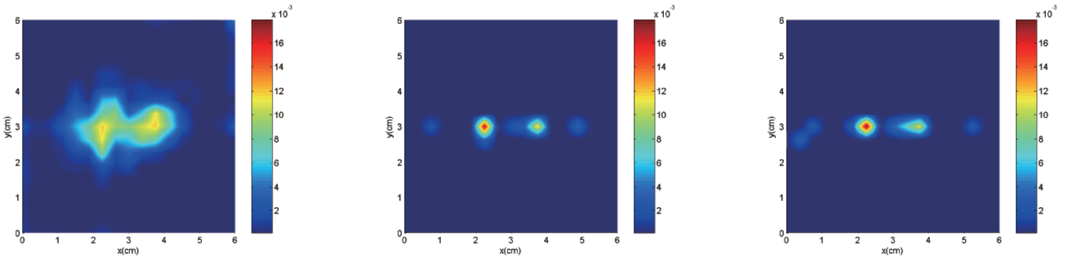
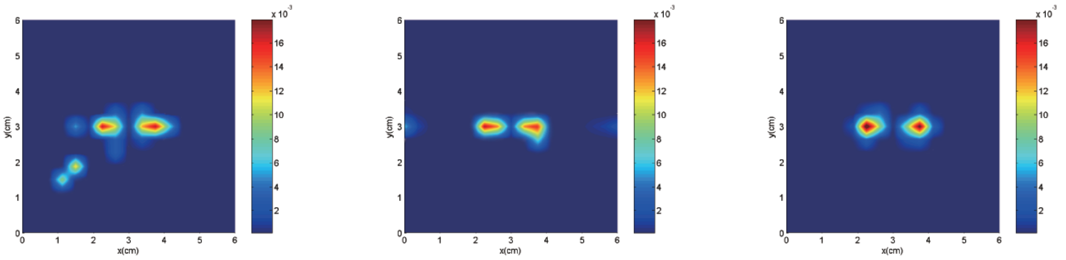


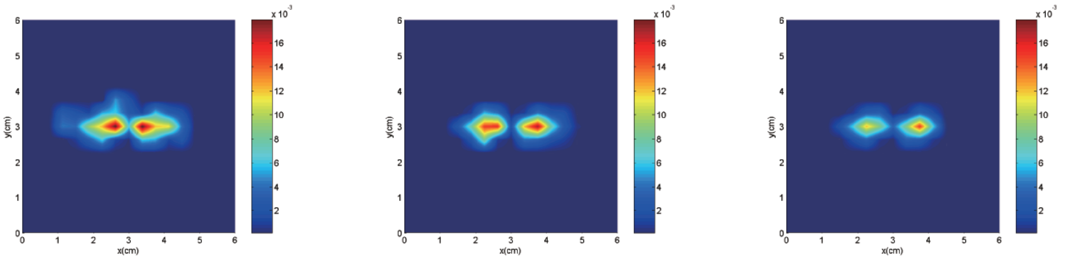
Figure 4.4: The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$ cm, $d = 1.5$ cm).



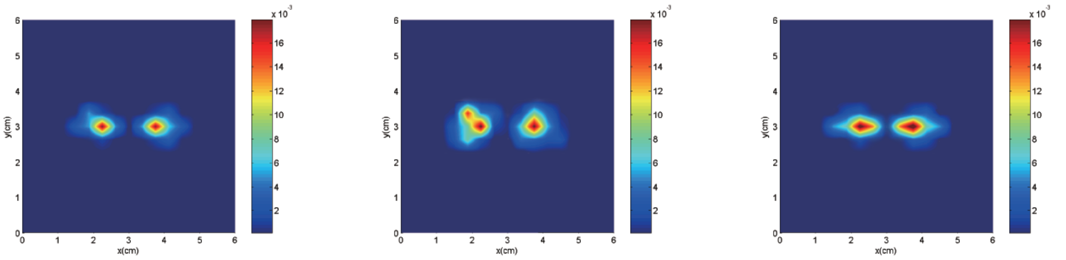
(a) Tikhonov (b) OMP (c) GP



(d) StOMP (e) ROMP (f) CoSaMP

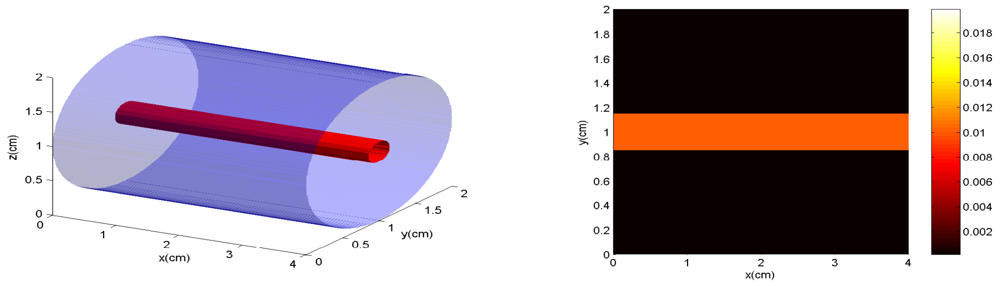


(g) BP-simplex (h) BP-interior (i) IST



(j) GPSR (k) FOCUSS (l_1) (l) FOCUSS (l_p)

Figure 4.5: The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using different algorithms, $r = 0.5$ cm, $d = 1.5$ cm, 1% noise.



(a) The silicon phantom

(b) The cross section of the fluorophore yield at $z = 1$ (middle) of the silicon phantom

Figure 4.6: The configuration of the real silicon phantom and the cross section of the fluorophore yield.

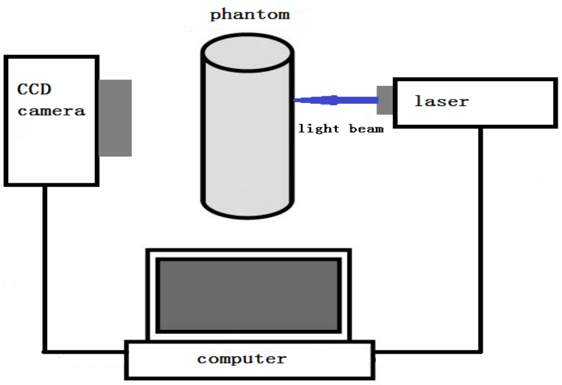
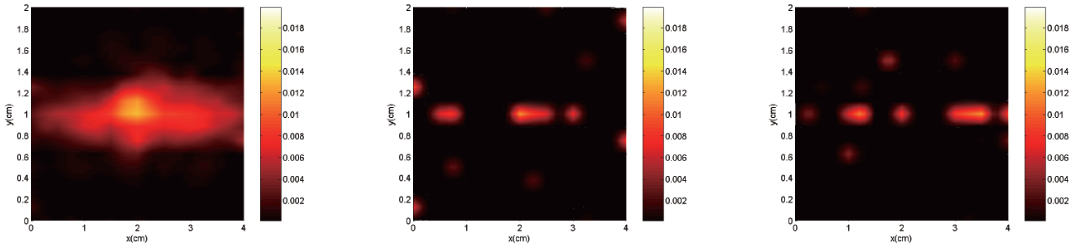
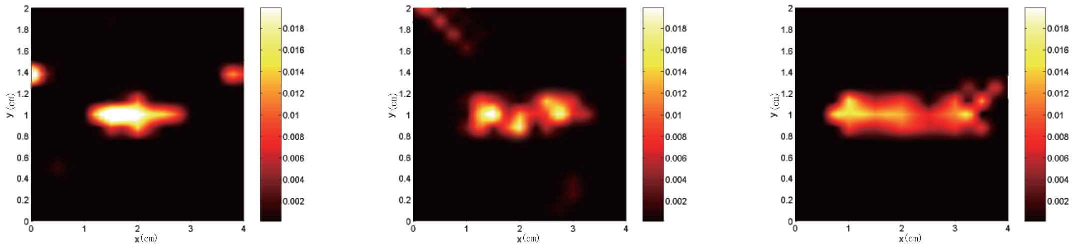


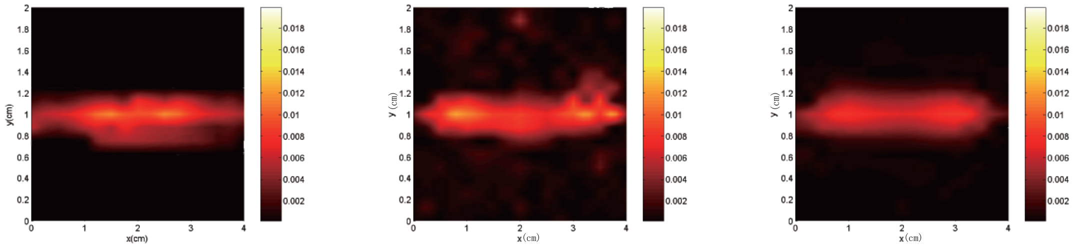
Figure 4.7: The imaging system used in the silicon phantom experiment.



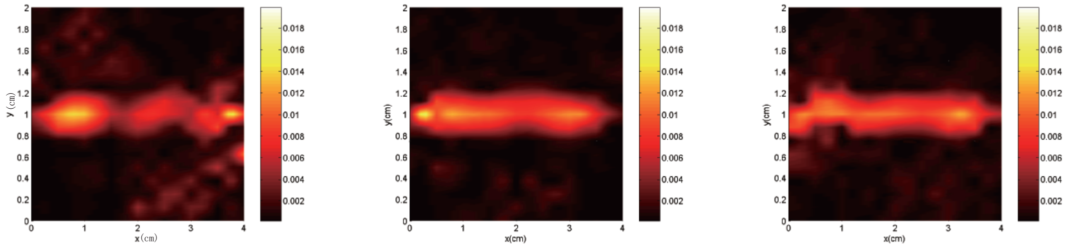
(a) Tikhonov (b) OMP (c) GP



(d) StOMP (e) ROMP (f) CoSaMP



(g) BP-simplex (h) BP-interior (i) IST



(j) GPSR (k) FOCUSS (Convex) (l) FOCUSS (Non-convex)

Figure 4.8: The cross sections of the phantom at $z = 1$ (middle) using different algorithms.

CHAPTER 5
FDOT RECONSTRUCTION WITH PRECONDITIONED FORWARD SENSING MATRIX

5.1 Introduction

In CS framework, sparse signals can be exactly recovered from limited number of measurements. In Chapter 4, we have reviewed a number of sparse signal recovery techniques in CS literature, and apply them in FDOT image reconstruction. A fundamental property that ensures accurate signal recovery is that the forward sensing matrix, which linearly projects the signal to the measure, is incoherent, i.e, the normalized inner products between two different columns of the sensing matrix are small [51–55]. However, in the FDOT inverse problem, although the fluorophore yield that we want to reconstruct is usually sparse in the imaging domain, the forward matrix, which maps the fluorophore yield to the boundary measurements, is often coherent. This is because of the diffusive nature of the optical domain, where the photon density of adjacent locations are similar due to mass scattering.

In many applications, the forward matrix can be designed to improve the recovery of sparse signals [52–54, 66, 77, 78]. Recent work has demonstrated that using a carefully designed preconditioning matrix can improve the performance of compressed sensing. In particular, a well-designed preconditioning matrix can reduce the coherence between the atoms of the equivalent dictionary, and as a consequence, reduce the reconstruction error.

In this chapter, we address the incoherence of the FDOT forward sensing matrix. We design a preconditioning matrix to reduce the coherence of the FDOT forward matrix. In 3D numerical simulation and real phantom experiment, we demonstrate the performance of the preconditioned sensing matrix in conjunction with the convex relaxation and greedy-type sparse signal recovery algorithms.

The rest of the chapter is organized as follows: Section 5.2 presents the preconditioned FDOT inverse problem formulation as a sparse signal recovery problem; Section 5.3 briefly reviews the coherence properties of the forward sensing matrix in

CS theory; In Section 5.4, we design the preconditioning matrix; In Sections 5.5 and 5.6, we present the performance of our approach in 3D numerical simulations and in a real silicon phantom experiment; Finally, Section 5.7 concludes this chapter.

5.2 Preconditioned FDOT Inverse Problem with Sparsity Constraint

Recall from Chapter 2 and Chapter 4, the linearized FDOT imaging problem with additive noise is

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}. \tag{5.1}$$

To recover sparse vector $\boldsymbol{x}$, we solve the l_0-norm or l_1-norm constraint optimization problem

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \text{ such that } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \varepsilon, \tag{5.2}$$

or

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ such that } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \varepsilon. \tag{5.3}$$

We design a preconditioning matrix to apply on $\boldsymbol{A}$ to improve the incoherence of the forward sensing matrix $\boldsymbol{A}$. Let $\boldsymbol{M}_A$ be the preconditioning matrix. When $\boldsymbol{M}_A$ is applied, (5.1) becomes

$$\boldsymbol{M}_A\boldsymbol{y} = \boldsymbol{M}_A\boldsymbol{A}\boldsymbol{x} + \boldsymbol{M}_A\boldsymbol{\epsilon} = \boldsymbol{A}_{pre} + \boldsymbol{M}_A\boldsymbol{\epsilon} \tag{5.4}$$

where $\boldsymbol{A}_{pre}$ denotes the preconditioned forward sensing matrix. Next, instead of solving (5.2) or (5.3), we solve the following minimization problem to reconstruct the fluorophore yield,

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \text{ such that } \|\boldsymbol{y}_{pre} - \boldsymbol{A}_{pre}\boldsymbol{x}\|_2 \leq \varepsilon_{pre}, \tag{5.5}$$

or

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ such that } \|\boldsymbol{y}_{pre} - \boldsymbol{A}_{pre}\boldsymbol{x}\|_2 \leq \varepsilon_{pre} \tag{5.6}$$

where $\boldsymbol{y}_{pre} = \boldsymbol{M}_A\boldsymbol{y}$ and $\|\boldsymbol{M}_A\boldsymbol{\epsilon}\|_2 \leq \varepsilon_{pre}$.

5.3 Performance Guarantees in Sparse Signal Reconstruction

In Chapter 4, we have reviewed the incoherence properties of the forward sensing matrix in CS literature. In this section, we recap the concept of mutual coherence and cumulative coherence relevant to our subsequent discussion.

In CS literature, many of the sufficiency conditions have been proposed to specify the incoherence of the forward matrix [51, 54, 55, 66, 77]. One of the most commonly used conditions is the mutual coherence [51],

$$\mathcal{M}(\boldsymbol{A}) = \max_{p,q,p\neq q} \frac{|\langle \boldsymbol{a}_p, \boldsymbol{a}_q\rangle|}{\|\boldsymbol{a}_p\|_2\|\boldsymbol{a}_q\|_2}. \tag{5.7}$$

$\mathcal{M}(\boldsymbol{A})$ is the largest normalized inner product of two different columns, which has a small value only if the columns of $\boldsymbol{A}$ are almost orthogonal to each other.

It was shown in [51, 55] that, if $\boldsymbol{x}_0$ is a vector that satisfies $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_0\|_2 \leq \boldsymbol{\varepsilon}$ with

$$\|\boldsymbol{x}_0\|_0 \leq \frac{1}{2}(1 + \frac{1}{\mathcal{M}(\boldsymbol{A})}), \tag{5.8}$$

then, $\boldsymbol{x}_0$ is necessarily the unique vector that satisfies $\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0\|_2 \leq \boldsymbol{\varepsilon}$ whose number of nonzero entries satisfies (5.8) [52, 55]. Furthermore, the difference between $\boldsymbol{x}_0$ and the solutions of the sparsity constraint optimization problems in (5.2) and (5.3) are only within a constant multiple of ε [52, 55].

The mutual coherence of $\boldsymbol{A}$ shows the worst-case analysis of the orthogonality of the columns of $\boldsymbol{A}$. To measure the average-case coherence of a matrix, the cumulative coherence or the Babel function has been proposed [77],

$$\mathcal{M}_1(k, \boldsymbol{A}) = \max_{p} \max_{|Q|=k, p\notin Q} \sum_{q\in Q} \frac{| < \boldsymbol{a}_p, \boldsymbol{a}_q > |}{\|\boldsymbol{a}_p\|_2\|\boldsymbol{a}_q\|_2}, \tag{5.9}$$

where Q is a subset of the columns in $\boldsymbol{A}$. It was shown that, if

$$\mathcal{M}_1(k, \boldsymbol{A}) + \mathcal{M}_1(k+1, \boldsymbol{A}) < 1, \tag{5.10}$$

then, the signal $\boldsymbol{x}_0$ with $\|\boldsymbol{x}_0\|_0 \leq k$ can be exactly recovered using greedy algorithms or basis pursuit [77]. Note that $\mathcal{M}_1(k, \boldsymbol{A})$ is an monotonically non-decreasing function in k, and $\mathcal{M}_1(1, \boldsymbol{A}) = \mathcal{M}(\boldsymbol{A})$. If $\mathcal{M}_1(k, \boldsymbol{A})$ increase slowly, $\boldsymbol{A}$ in said to be quasi-incoherent.

More generally, a matrix with N columns have $\frac{N(N-1)}{2}$ normalized inner products between different pairs of columns. It was observed that as the average value of the normalized inner products between different columns decreases, the performance of sparse signal recovery methods typically improves [52, 53, 66, 77]. In particular, it was shown that a well-designed preconditioning matrix can reduce the normalized inner products between the columns of the sensing matrix, and as a consequence, reduce the mean square error of the reconstructed images [52, 53, 124, 133].

5.4 Preconditioning Matrix Design

A majority of works on CS design a random preconditioning matrix with an orthogonal dictionary, such that the resultant sensing matrix $\boldsymbol{A}_{pre}$ simply satisfies the incoherence or RIP property with high probability. However, in the case where the dictionary is overcomplete and thus unorthogonal, a random matrix does not lead to optimal CS property. Recently, some researches have shown that a carefully designed deterministic projection matrix achieved better incoherence property of the sensing matrix [52, 53, 124, 133]. To minimize the coherence of $\boldsymbol{A}_{pre}$, we seek to determine $\boldsymbol{M}_A$ such that the Gramm type matrix $\tilde{\boldsymbol{A}}_{pre}^T \tilde{\boldsymbol{A}}_{pre}$ approximates the identity matrix,

$$\boldsymbol{G}_A = \tilde{\boldsymbol{A}}_{pre}^T \tilde{\boldsymbol{A}}_{pre} \approx \boldsymbol{I} \tag{5.11}$$

where $\tilde{\boldsymbol{A}}_{pre}$ is an equivalent form of $\boldsymbol{A}_{pre}$ with all the columns normalized.

Without loss of generality for the rest of this chapter, we assume that the columns of $\boldsymbol{A}$ and $\boldsymbol{A}_{pre}$ are normalized to unity and drop tilde to simplify our notation for rest of the chapter.

5.4.1 Elad's Method

In [52], the authors proposed to minimize its t-averaged mutual coherence, which was defined as the average of all absolute and normalized inner products between different columns in $\boldsymbol{A}$ above some fixed threshold value t,

$$\mathcal{M}_t(\boldsymbol{A}) = \frac{\sum_{p,q,p\neq q}(|g_{A_{p,q}}| > t) \cdot |g_{A_{p,q}}|}{\sum_{p,q,p\neq q}(|g_{A_{p,q}}| > t)}. \tag{5.12}$$

The object is to reduce the off diagonal entries of $\boldsymbol{G}_A$ with an absolute value larger than t, which in turns reduces $\mu_t(\boldsymbol{A})$. The reduction of $\mu_t(\boldsymbol{A})$ was done iteratively. In each iteration, $g_{A_{p,q}} > t$ is shrinked by a small amount using the following function:

$$g_{A_{p,q}} = \begin{cases} \gamma g_{A_{p,q}} & |g_{A_{p,q}}| > t \\ \gamma t sign(g_{A_{p,q}}) & \gamma t \leq |g_{A_{p,q}}| \leq t \\ g_{A_{p,q}} & |g_{A_{p,q}}| < \gamma t \end{cases}. \tag{5.13}$$

The next steps forces the rank to be M and find the matrix $\boldsymbol{M}_A$ that could best fit the new shrinked Gram-type matrix. The process could be realized using SVD. The details can be found in [52].

5.4.2 Duarte-Carvajalino and Sapiro's Method

In [53], the authors proposed an alternative algorithm to design the preconditioning matrix $\boldsymbol{M}_A$. Instead of reducing $\mu_t(\boldsymbol{A})$, the authors sought the solution of the following optimization problem,

$$\min_{\boldsymbol{M}_A} \|\boldsymbol{A}_{pre}^T\boldsymbol{A}_{pre} - \boldsymbol{I}\|_F = \|(\boldsymbol{M}_A\boldsymbol{A})^T\boldsymbol{M}_A\boldsymbol{A} - \boldsymbol{I}\|_F, \tag{5.14}$$

where the subscription F denotes the Frobenius norm.

Multiplying both sides of (5.11) with $\boldsymbol{A}$ on the left and $\boldsymbol{A}^T$ on the right, it becomes

$$\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{M}_A^T\boldsymbol{M}_A\boldsymbol{A}\boldsymbol{A}^T \approx \boldsymbol{A}\boldsymbol{A}^T. \tag{5.15}$$

Consider the eigenvalue decomposition $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, (5.15) is equivalent to

$$\boldsymbol{\Lambda}\boldsymbol{U}^T\boldsymbol{M}_A^T\boldsymbol{M}_A\boldsymbol{U}\boldsymbol{\Lambda} \approx \boldsymbol{\Lambda}. \tag{5.16}$$

Thus, (5.14) can be reduced to the following optimization problem

$$\min_{\boldsymbol{M}_A} \|\boldsymbol{\Lambda} \boldsymbol{U}^T \boldsymbol{M}_A^T \boldsymbol{M}_A \boldsymbol{U} \boldsymbol{\Lambda} - \Lambda\|_F \tag{5.17}$$

The details can be found in [53].

The authors also compared their work with the method described [52] to show that their algorithm was more computationally efficient, and the learned projection matrix lead to better incoherence property of the sensing matrix $\boldsymbol{A}_{pre}$.

5.4.3 Schnass and Vandergheynst's Method

In [124], the authors defined the cumulative cross-coherence or cross Babel function of original and the preconditioned forward matrix as follows,

$$\mathcal{M}_1(k, \boldsymbol{A}_{pre}, \boldsymbol{A}) = \max_p \max_{|Q|=k, p \notin Q} \frac{\boldsymbol{a}_p^T \boldsymbol{a}_{pre_q}}{\|\boldsymbol{a}_p\|_2 \|\boldsymbol{a}_{pre_q}\|_2}. \tag{5.18}$$

The author tried to find $\boldsymbol{A}_{pre}$ that minimize $\mathcal{M}_1$ in (5.18). This is equivalent as finding $\boldsymbol{A}_{pre}$ such that the Gram-type matrix $\boldsymbol{G}' = \boldsymbol{A}_{pre}\boldsymbol{A}$ is close to identity matrix. Two sets of matrices were defined as follows:

$$\begin{aligned} \mathcal{G} &:= \{\boldsymbol{G}' = \boldsymbol{A}_{pre}\boldsymbol{A}\} \\ \mathcal{H} &:= \{\boldsymbol{H}, \text{an } N \times N \text{ matrix with } H_{p,p} = 1 \text{ and } |H_{p,q} < t \text{ for } p \neq q\}. \end{aligned} \tag{5.19}$$

$\boldsymbol{A}_{pre}$ can be found by solving the following optimization problem iteratively,

$$\min_{\boldsymbol{A}_{pre}} \|\boldsymbol{G}' - \boldsymbol{H}\|_F \quad \text{such that} \quad \boldsymbol{G}' \in \mathcal{G}, \boldsymbol{H} \in \mathcal{H}, \tag{5.20}$$

see [124] for details.

5.4.4 Our Approach

In this chapter, we use an approach similar to the one described in [53] to precondition the FDOT forward sensing matrix. We look for $\boldsymbol{M}_A$ that solves the optimization problem in (5.14), and consider multiplying both sides of $\boldsymbol{A}_{pre}^T \boldsymbol{A}_{pre} \approx \boldsymbol{I}$

with $\boldsymbol{A}$ on the left and $\boldsymbol{A}^T$ on the right,

$$\boldsymbol{A}\boldsymbol{A}_{pre}^T\boldsymbol{A}_{pre}\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{M}_A^T\boldsymbol{M}_A\boldsymbol{A}\boldsymbol{A}^T \approx \boldsymbol{A}\boldsymbol{A}^T. \tag{5.21}$$

Let

$$\boldsymbol{A} = \boldsymbol{U}_A\boldsymbol{\Sigma}_A\boldsymbol{V}_A^T \tag{5.22}$$

be the singular value decomposition of $\boldsymbol{A}$, and substitute (5.22) into (5.21), we obtain

$$\boldsymbol{U}_A\boldsymbol{\Sigma}_A\boldsymbol{V}_A^T\boldsymbol{V}_A\boldsymbol{\Sigma}_A^T\boldsymbol{U}_A^T\boldsymbol{M}_A^T\boldsymbol{M}_A\boldsymbol{U}_A\boldsymbol{\Sigma}_A\boldsymbol{V}_A^T\boldsymbol{V}_A\boldsymbol{\Sigma}_A^T\boldsymbol{U}_A^T \approx \boldsymbol{U}_A\boldsymbol{\Sigma}_A\boldsymbol{V}_A^T\boldsymbol{V}_A\boldsymbol{\Sigma}_A^T\boldsymbol{U}_A^T \tag{5.23}$$

(5.23) can be simplified as follows,

$$(\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_A^T)\boldsymbol{U}_A^T\boldsymbol{M}_A^T\boldsymbol{M}_A\boldsymbol{U}_A(\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_A^T) \approx (\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_A^T) \tag{5.24}$$

Thus, we choose $\boldsymbol{M}_A$ as

$$\boldsymbol{M}_A = (\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_A^T)^{-1/2}\boldsymbol{U}_A^T. \tag{5.25}$$

In practice, $\boldsymbol{A}$ is usually ill-conditioned with a large number of singular values equal or close to 0. Therefore, $\boldsymbol{M}_A$ in (6.34) is also ill-conditioned with a condition number approaching to infinity. To mitigate this, we regularize (6.34) and write:

$$\boldsymbol{M}_A = (\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_A^T + \lambda\boldsymbol{I})^{-1/2}\boldsymbol{U}_A^T \tag{5.26}$$

where λ is small a regularization constant.

5.5 Numerical Simulations

In this section, we present numerical simulations to demonstrate the effects of preconditioning matrix on the coherence of the FDOT forward sensing matrix as well as on the quality of the reconstructed FDOT images.

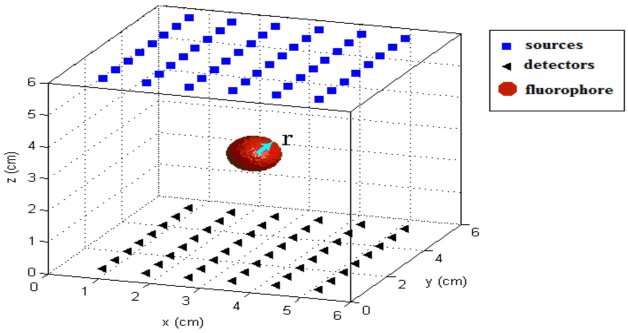


Figure 5.1: An illustration of the phantom and source-detector configuration.

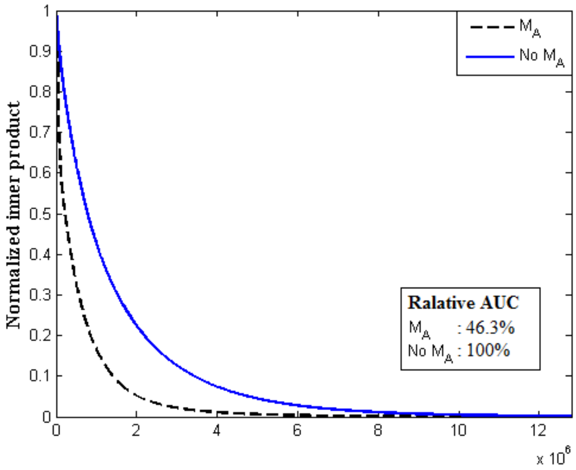
5.5.1 Simulation Setup

We set up a $6 \times 6 \times 6$ cm^3 cubic phantom shown in Figure 5.1. We set the background absorption coefficient to $\mu_{am} = 0.05$ cm^{-1}, and the diffusion coefficient to $D(\boldsymbol{r}) = 0.04$ cm^{-1} at both the excitation and emission wavelengths. The circular heterogeneity with radius r simulated the fluorophore concentration with fluorophore absorption coefficient $\mu_{axf} = 0.005$ cm^{-1}. 48 sources (shown in squares) and 48 detectors (shown in triangles) are uniformly placed at the top and bottom of the imaging domain. We discretized the imaging domain into $20 \times 20 \times 20$ voxels. Thus, the forward sensing matrix is 2304 by 8000. We simulated different data sets corresponding to 3 different values of the radius r : 0.5, 0.75, and 1.0 cm.

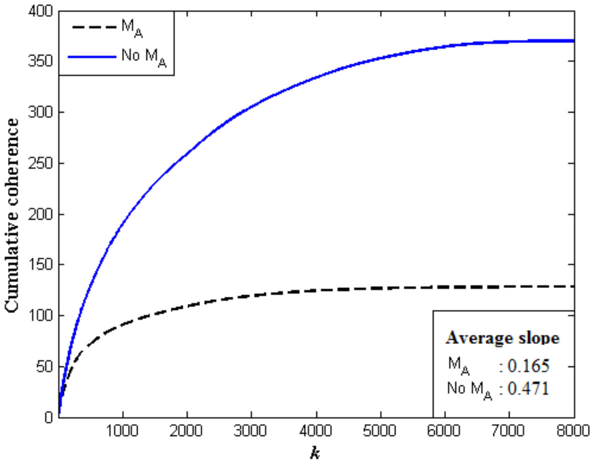
To simulate the measurement noise, we assumed that, when a sufficiently large number of photons are detected, the noise at each detector can be approximated by a Gaussian random variable with its variance proportional to the magnitude of the detector reading. The SNR of the measurements is defined as

$$SNR = 20 \log_{10} \frac{\|\boldsymbol{\Gamma}\|_2}{\|\boldsymbol{\epsilon}\|_2}. \tag{5.27}$$

We simulated three sets of noise contaminated measurements with approximately 10%, 3% and 1% noise, corresponding to the SNR value of 20 dB, 30 dB and 40 dB, respectively.



(a) Top 40% values of all the normalized inner products



(b) Cumulative coherence

Figure 5.2: The normalized inner products and cumulative coherence of the forward matrix before and after applying the preconditioning matrix.

5.5.2 Simulation Results - Coherence of the FDOT Forward Matrix

We evaluated the coherence of the forward matrix before and after the application of the preconditioning matrix. Figure 5.2(a) shows the largest 40% of all the normalized inner products between different columns of the forward sensing matrix, arranged in a descending order. The remaining 60% of the normalized inner products between different columns are close to 0. The application of preconditioning matrix reduces the large correlations between different columns of $\boldsymbol{A}$. To quantify the improvement, we computed the *area under the curves* (AUC) for each case, and the relative percentages of AUC with respect to the no-mask case. The results of relative AUC are given in the box in Figure 5.2(a). We see that the preconditioning matrix reduces the AUC to 46.3%.

Figure 5.2(b) shows the plot of the cumulative coherence $\mathcal{M}_1(k, \boldsymbol{A})$ as a function of k. As stated before, the cumulative coherence is a nondecreasing function of k. When $\mathcal{M}_1(k, \boldsymbol{A})$ increases slowly, the forward matrix is said to be quasi-incoherent [77]. From Figure 5.2(b), when preconditioning matrix is applied to the light sources and detectors, the cumulative coherence increases much slower as a function of k than that of no-mask case. The average slopes of the curves are shown in the box in Figure 5.2(b) for a quantitative comparison.

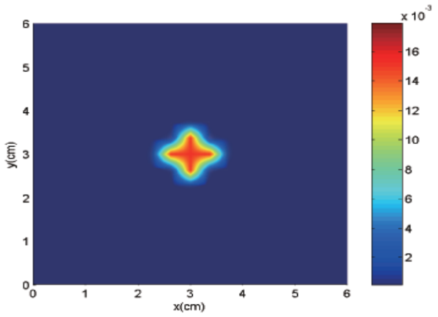
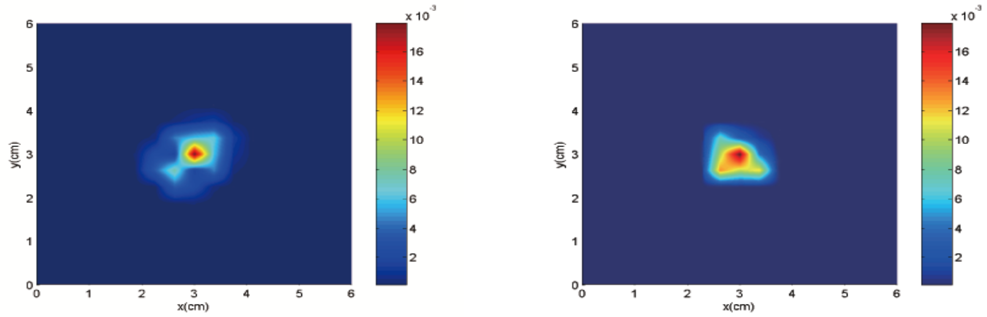


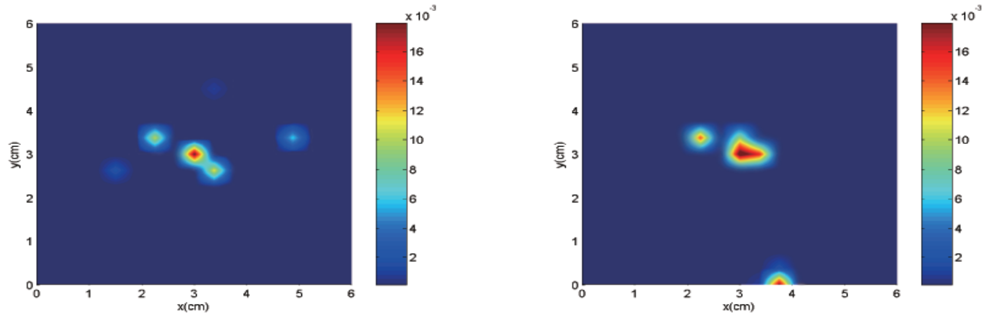
Figure 5.3: The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$ cm).

5.5.3 Simulation Results - Image Reconstruction

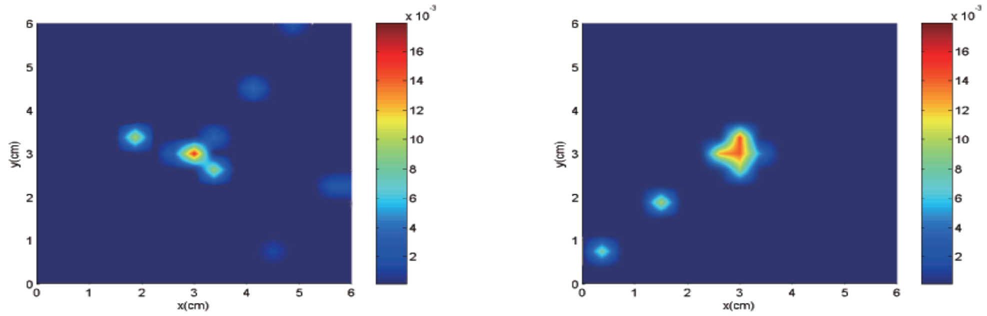
In this simulation, we reconstructed the fluorophore concentration map at 3 different sparsity levels corresponding to 3 different radii using measurements at 3 different noise levels. In each case, we used the forward matrix with and without application of the preconditioning matrix. We used six different sparsity promoting reconstruction methods available in the CS literature (See Chapter 4 for more details). Specifically, we used: StOMP [62], ROMP [63], CoSaMP [65], BP-interior [67], IST [69], and GPSR [71]. The first three are greedy type algorithms, and the last three are convex relaxation based algorithms.



(a) StOMP (no $\boldsymbol{M}_A$) (b) StOMP ($\boldsymbol{M}_A$)



(c) ROMP (no $\boldsymbol{M}_A$) (d) ROMP ($\boldsymbol{M}_A$)



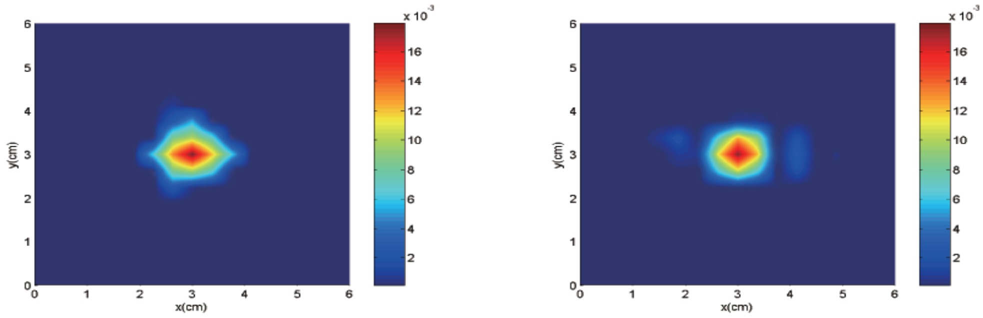
(e) CoSaMP (no $\boldsymbol{M}_A$) (f) CoSaMP ($\boldsymbol{M}_A$)

Figure 5.4: The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using greedy algorithms, $r = 0.5$ cm, 1% noise.

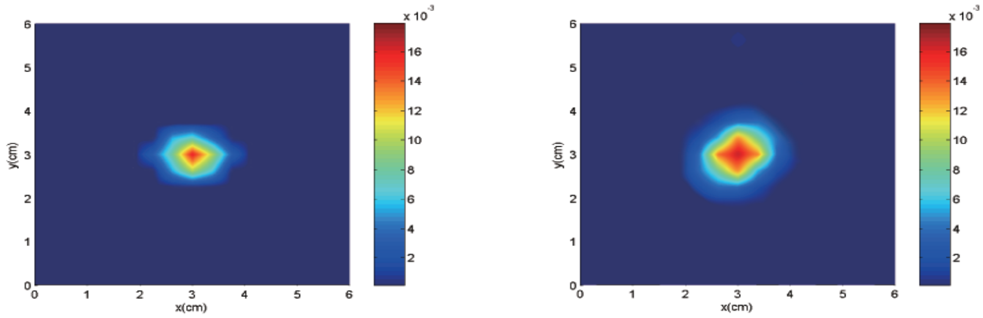
We used NMSE to measure the accuracy of the reconstructed images,

$$\text{NMSE} = \frac{\|\boldsymbol{x}_{\text{true}} - \boldsymbol{x}_{\text{recon}}\|_2}{\#\text{ of voxels}}. \tag{5.28}$$

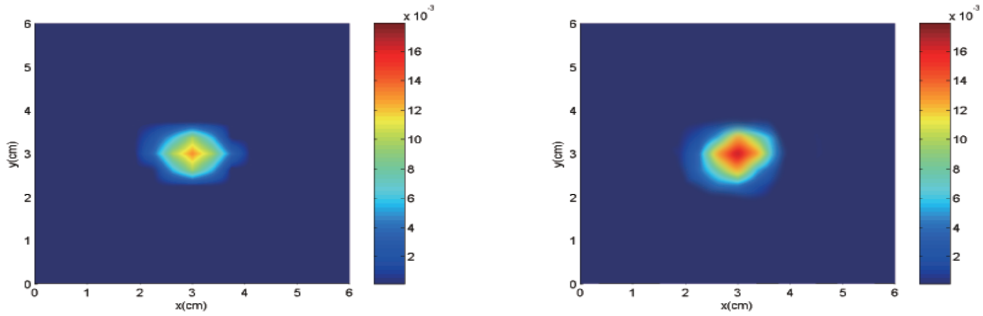
We averaged the NMSE of the reconstructed images over 30 realizations of noise. The results are tabulated in Table 5.1. The results show that, in general, the NMSE of the reconstructed images increases as the sparsity level of the signal increases. By



(a) BP-interior (no $\boldsymbol{M}_A$) (b) BP-interior ($\boldsymbol{M}_A$)



(c) IST (no $\boldsymbol{M}_A$) (d) IST ($\boldsymbol{M}_A$)



(e) GPSR (no $\boldsymbol{M}_A$) (f) GPSR ($\boldsymbol{M}_A$)

Figure 5.5: The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using convex relaxation algorithms, $r = 0.5$ cm, 1% noise.

applying the preconditioning matrix $\boldsymbol{M}_A$, the NMSE reduced for all reconstructions. Table 5.1 shows a greater reduction in NMSE for greedy-type algorithms than that of convex relaxation techniques.

Figure 5.3 shows the cross sections of the reconstructed images at the middle of the imaging domain when the radius of the heterogeneity is $r = 0.5$ cm. The cross sections of the reconstructed fluorophore yield maps using greedy type algorithms

Table 5.1: NMSE of the reconstructed images (10^{-4}).

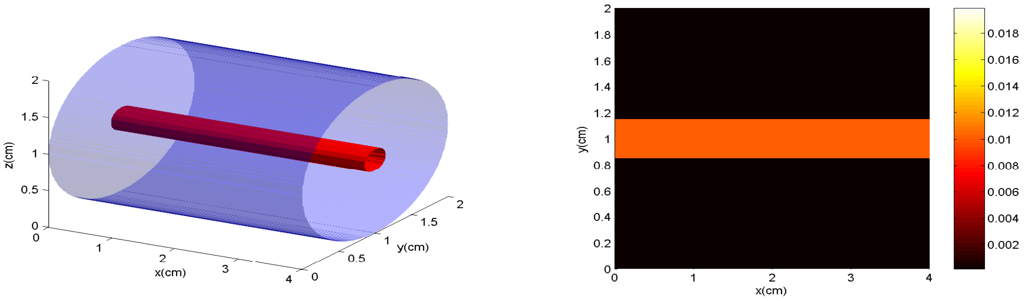
Algorithm		$r = 0.5$ cm			$r = 0.75$ cm			$r = 1.0$ cm		
StOMP	SNR (dB)	40	30	20	40	30	20	40	30	20
	No $\boldsymbol{M}_A$	2.05	6.47	12.1	3.87	14.5	28.0	9.44	31.8	63.7
	$\boldsymbol{M}_A$	1.40	4.20	8.51	2.74	9.65	20.6	6.94	22.2	45.2
ROMP	SNR (dB)	40	30	20	40	30	20	40	30	20
	No $\boldsymbol{M}_A$	1.66	5.85	13.2	3.68	13.0	31.6	8.01	29.0	64.2
	$\boldsymbol{M}_A$	1.26	3.76	8.13	2.09	11.8	16.4	5.77	22.4	38.5
CoSaMP	SNR (dB)	40	30	20	40	30	20	40	30	20
	No $\boldsymbol{M}_A$	2.11	6.35	13.0	5.97	14.5	28.9	11.8	34.7	67.8
	$\boldsymbol{M}_A$	1.24	3.28	8.18	3.78	9.73	22.0	6.26	23.7	49.1
BP-interior	SNR (dB)	40	30	20	40	30	20	40	30	20
	No $\boldsymbol{M}_A$	1.56	3.03	7.05	2.44	7.13	15.3	5.03	20.8	39.1
	$\boldsymbol{M}_A$	1.04	2.07	5.96	1.87	6.09	13.2	4.31	18.7	37.2
IST	SNR (dB)	40	30	20	40	30	20	40	30	20
	No $\boldsymbol{M}_A$	1.20	3.55	7.49	3.51	6.97	19.2	6.38	23.3	49.9
	$\boldsymbol{M}_a$	1.07	2.98	4.74	2.43	5.50	11.8	5.80	20.7	45.1
GPSR	SNR (dB)	40	30	20	40	30	20	40	30	20
	No $\boldsymbol{M}_A$	1.29	3.28	7.07	3.41	7.13	17.1	5.51	22.2	46.6
	$\boldsymbol{M}_A$	1.06	2.24	6.14	2.10	5.67	13.5	4.40	16.4	39.4

are shown in Figure 5.4, and those using convex relaxation techniques are shown in Figure 5.5. For each type of algorithm, the application of the preconditioning matrix $\boldsymbol{M}_A$ leads to better reconstruction. The visual improvements are more obvious for the greedy algorithms than the relaxation algorithms. In the greedy type algorithms, the support of the signal is determined by selecting the columns of the forward matrix that have the greatest correlation with the measurements. The reduction of the normalized inner products between different columns in the forward matrix has a direct effect on the column selection procedure at each iteration.

5.6 Silicon Phantom Experiment

5.6.1 Real Phantom Configuration

Figure 5.6(a) shows an illustration of the cylindrical phantom used in the experiment. The phantom was made of silicone rubber with diameter of about 2 cm, and length of 4 cm. The phantom had homogeneous absorption coefficient $\mu_a = 0.2$ cm^{-1} and scattering coefficient $\mu_s' = 12$ cm^{-1} ($D = 1/3(\mu_a + \mu_s')$) at



(a) The silicon phantom

(b) The cross section of the fluorophore yield at $z = 1$ (middle) of the silicon phantom

Figure 5.6: The configuration of the real silicon phantom and the cross section of the fluorophore yield.

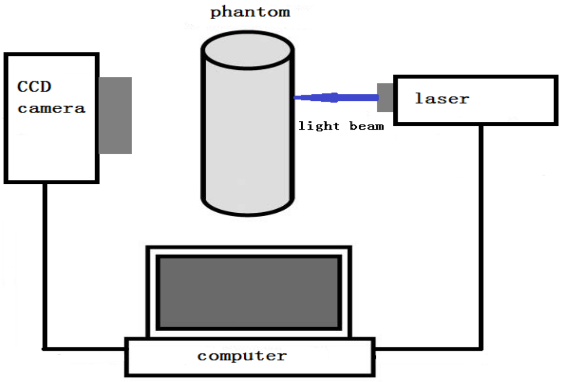


Figure 5.7: The imaging system used in the silicon phantom experiment.

both the excitation and emission wavelengths (743 nm and 767 nm). The silicon phantom contained a hollow cylindrical tube in the middle with approximately 3 mm in diameter. which was filled with intralipid and ink to mimic the same optical properties as the background. The intralipid and ink contained 1 micromolar of Cy7 as the fluorophore. The cross section of the fluorophore yield at $z = 1$ is shown in Figure 5.6(b).

The FDOT measurements were collected using the FDOT imaging system reported in [132], as illustrated in Figure 5.7. Specifically, focused collimated laser beams were used as point light sources to excite the fluorophore. We had 60 point sources in total. The fluorescence measurements were collected by an electrically cooled CCD camera. The reading of the detector was recorded as the mean value of a subregion with 5×5 pixels around each detector location. We selected 60 detector locations. We discretized the imaging domain into $20 \times 20 \times 20$ voxels. Thus, the forward sensing matrix is of dimension 3600 by 8000.

5.6.2 Coherence of the FDOT Forward Matrix

We evaluated the coherence of the forward matrix with and without the preconditioning matrix $\boldsymbol{M}_A$. Figure 5.8(a) shows the largest 40% of all the normalized inner products between different columns of the forward sensing matrix, arranged in a descending order. The remaining 60% of the normalized inner products between different columns are close to 0. Clearly, application of $\boldsymbol{M}_A$ reduces the large correlations between different columns of $\boldsymbol{A}$. To quantify the improvement, we computed the relative *area under the curves* (AUC), which is given in the box in Figure 5.8(a). We see that the preconditioning matrix reduces the AUC to 39.59%. Figure 5.8(b) shows the plot of the cumulative coherence. When $\boldsymbol{M}_A$ is applied, the cumulative coherence increases much slower. The average slope of each curve is provided in the box in Figure 5.8(b).

The cross sections of the reconstructed fluorophore yield maps using greedy-type algorithms are shown in Figure 5.9, and those using convex relaxation techniques are shown in Figure 5.10. We observe that the application of the preconditioning matrix results in reconstructed images that are in better agreement with the original fluorophore yield map. However, there exists some background noise in the reconstructed images. This is due to the large condition number of $\boldsymbol{M}_A$ which amplifies the measurement noise. The visual improvements are most obvious for the greedy-type algorithms. In simple greedy-type algorithms, the support of the signal is determined by selecting the columns of the forward matrix that have the greatest correlation with the measurements. The reduction of the coherence in the forward matrix has a direct effect on the column selection procedure at each iteration.

To quantitatively assess the sparse reconstruction results, we calculated the *contrast to background noise ratio* (CBNR) of the reconstructed fluorophore yield map. CBNR is defined as the ratio between the mean value of the foreground fluorophore region and the standard deviation of the background. The results,

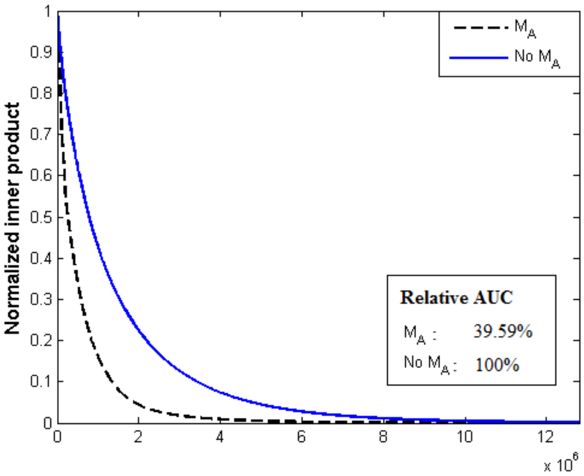
Table 5.2: CBNR of the reconstructed fluorophore images using different algorithms.

Greedy	StOMP	ROMP	CoSaMP
$\boldsymbol{M}_A$	20.9	28.6	32.3
No $\boldsymbol{M}_A$	16.5	18.8	26.7
Relaxation	BP-interior	IST	GPSR
$\boldsymbol{M}_A$	23.8	19.8	22.1
No $\boldsymbol{M}_A$	14.9	17.1	12.3

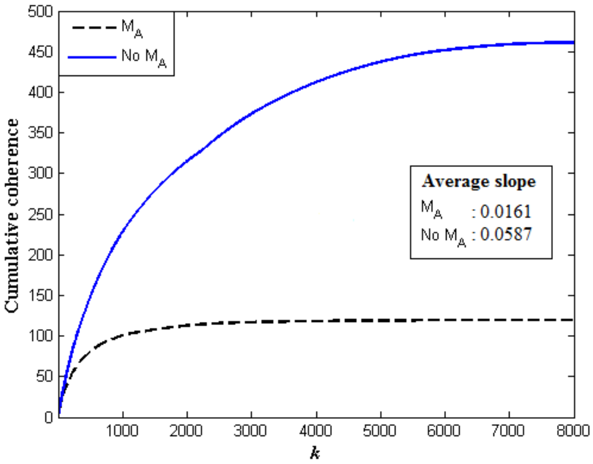
summarized in Table 5.2, indicate the improvements in the image contrasts due to application of the preconditioning matrix using different sparse signal recovery algorithms.

5.7 Conclusion

In this chapter, we demonstrated the application of the preconditioning matrix in sparse FDOT reconstruction based on CS framework. In CS theory, accurate recovery of a sparse signal from an underdetermined linear system requires the underlying forward matrix to be incoherent. We showed that the application of the preconditioning matrix is able to reduce the coherence of the FDOT forward sensing matrix. To reconstruct the fluorophore concentration map, we used sparse signal recovery techniques, including both the greedy type and convex relaxation algorithms. We showed that the application of the preconditioning matrix in conjunction with sparse signal recovery techniques improves the visual quality of reconstructed images and reduces the mean square error in both numerical simulations and in a real phantom experiment.

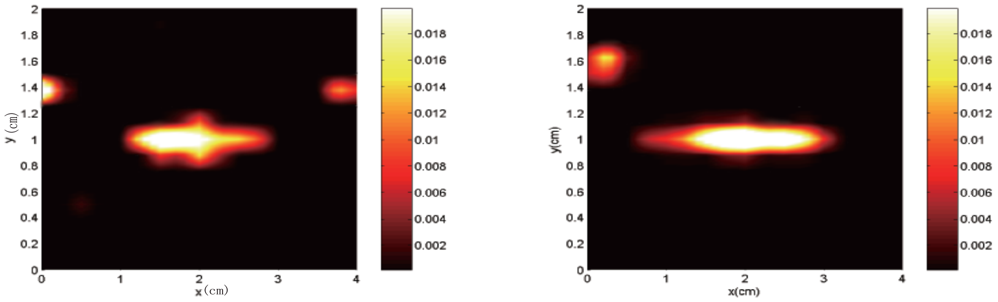


(a) Top 40% values of all the normalized inner products

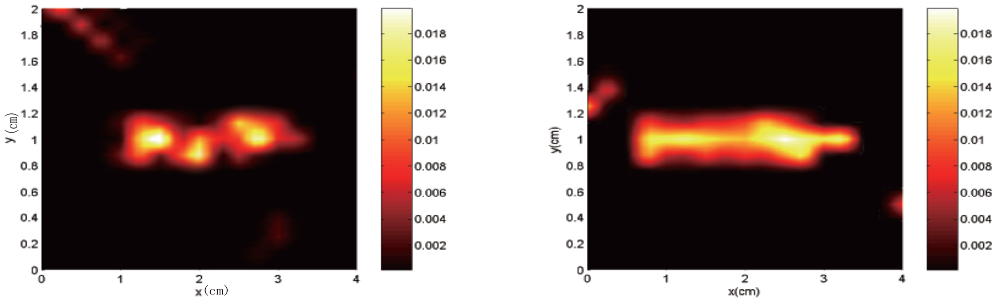


(b) Cumulative coherence

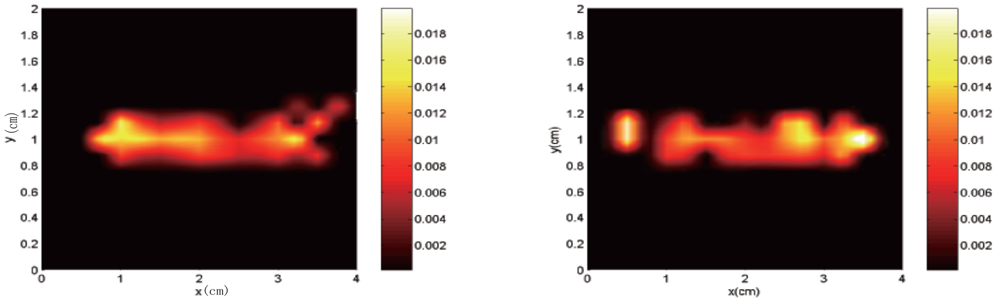
Figure 5.8: The normalized inner products and cumulative coherence of the forward matrix before and after applying the preconditioning matrix.



(a) StOMP (no $\boldsymbol{M}_A$) (b) StOMP ($\boldsymbol{M}_A$)

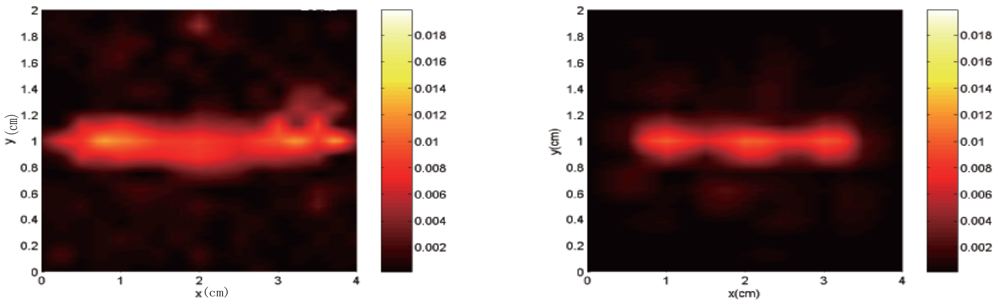


(c) ROMP (no $\boldsymbol{M}_A$) (d) ROMP ($\boldsymbol{M}_A$)

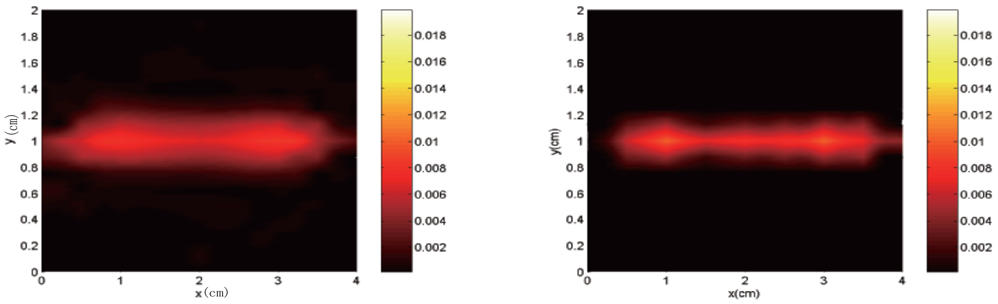


(e) CoSaMP (no $\boldsymbol{M}_A$) (f) CoSaMP ($\boldsymbol{M}_A$)

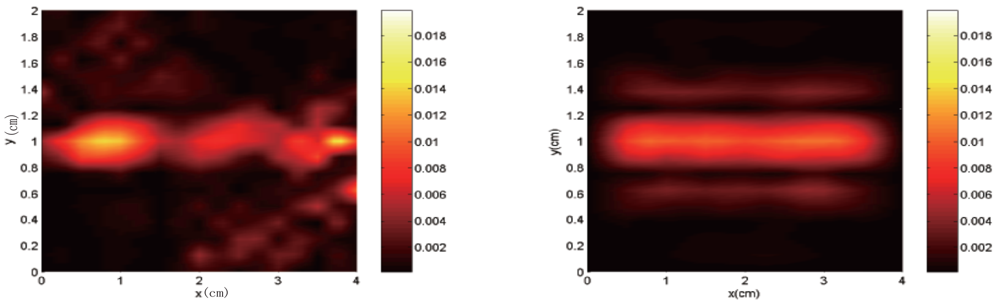
Figure 5.9: The cross sections at $z = 1$ (middle) of the reconstructed phantom using greedy algorithms.



(a) BP-interior (no $\boldsymbol{M}_A$) (b) BP-interior ($\boldsymbol{M}_A$)



(c) IST (no $\boldsymbol{M}_A$) (d) IST ($\boldsymbol{M}_A$)



(e) GPSR (no $\boldsymbol{M}_A$) (f) GPSR ($\boldsymbol{M}_A$)

Figure 5.10: The cross sections at $z = 1$ (middle) of the phantom using convex relaxation techniques.

CHAPTER 6
LIGHT ILLUMINATION AND DETECTION PATTERNS FOR FDOT BASED ON COMPRESSIVE SENSING

6.1 Introduction

In Chapter 5, we have applied a preconditioning matrix to improve the incoherence of the FDOT forward sensing matrix $\boldsymbol{A}$. In this chapter, we use an alternative approach to improve the incoherence of $\boldsymbol{A}$ by designing the light illumination and detection patterns.

In FDOT, the imaging domain is often illuminated sequentially by a number of point sources placed at the boundary of the imaging domain. However, recent work has demonstrated that the "structured illumination" can be useful in reducing the ill-posedness of the optical tomography inverse problem, and to improve the spatial resolution of the reconstructed images [79–86].

In this chapter, unlike the conventional FDOT imaging where point sources with constant intensity are time-multiplexed, i.e., turned on one at a time, we assume that multiple spatially distributed light sources illuminate the medium simultaneously and the corresponding boundary measurements are linearly filtered prior to image reconstruction. We show that the FDOT forward matrix can be expressed as a columnwise Kronecker product of two matrices. The first one is determined by the excitation light field which depends on the spatial configuration of light sources and their intensities as well as the Green's function of the endogenous background. The second matrix is determined by the emission light field which depends on the spatial configuration of the detectors and the Green's function of the emission light field. The incoherence of the FDOT forward sensing matrix is related to the incoherence of these two underlying matrices. We design two preconditioners, which we refer to as optical and measurement masks, to reduce the coherence of these two matrices. The optical mask is an intensity matrix, with each of its row being an intensity/illumination pattern applied to the point sources. The measurement masks is a linear filter applied to the measured data collected by the detectors in each illumination.

We present extensive numerical simulations and a real silicon phantom experiment to show that the application of optical and measurement masks reduces the coherence of the FDOT forward matrix. We use a number of sparse signal recovery techniques, including greedy type and convex relaxation algorithms to demonstrate the improvements in fluorophore reconstruction when optical and measurement masks are applied.

Our approach can be used to determine not only the optimal source intensities, but also the location of sources and detectors. Given a fluorophore designed to accumulate in certain regions and anatomical a priori information, optimal source detector locations and source intensities can be determined via numerical simulations assuming a large array of sources and detectors prior to the imaging process. This may eliminate unnecessary illumination, optimize imaging process and result in better image quality than the conventional FDOT imaging process.

The rest of this chapter is organized as follows: In Section 6.2, we derive bounds on the incoherence of the FDOT forward matrix; In 6.3, we present the design of the optical and measurement masks; In Sections 6.4 and 6.5, we present the performance of our approach in 3D numerical simulations and in a real silicon phantom experiment; Finally, Section 6.6 concludes the chapter.

6.2 Bounds on the Coherence of the FDOT Forward Matrix

In this section, we present the inherent structure of the FDOT forward matrix and show that it is the column-wise Kronecker product of two matrices: A matrix that is composed of the discretized excitation light fields; and another matrix composed of the discretized Green's function of the diffusion equation governing the emission light field propagation. We show that the normalized inner product between two different columns of the FDOT forward matrix can be expressed as a product of the normalized inner products between the columns of these two underlying matrices. We next derive upper bounds on the mutual coherence and cumulative coherence of the FDOT forward matrix. These observations are essential in the design and optimization of the FDOT forward matrix.

Recall that the FDOT forward matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \begin{bmatrix} g^1_{em,1}\phi^1_{ex,1} & g^1_{em,2}\phi^1_{ex,2} & \cdots & g^1_{em,N}\phi^1_{ex,N} \\ \vdots & & & \vdots \\ g^{N_d}_{em,1}\phi^1_{ex,1} & g^{N_d}_{em,2}\phi^1_{ex,2} & \cdots & g^{N_d}_{em,N}\phi^1_{ex,N} \\ g^1_{em,1}\phi^2_{ex,1} & g^1_{em,2}\phi^2_{ex,2} & \cdots & g^1_{em,N}\phi^2_{ex,N} \\ \vdots & & & \vdots \\ g^{N_d}_{em,1}\phi^{N_s}_{ex,1} & g^{N_d}_{em,2}\phi^{N_s}_{ex,2} & \cdots & g^{N_d}_{em,N}\phi^{N_s}_{ex,N} \end{bmatrix}.$$

We see that the entries of the $\boldsymbol{A}$ is the product of two quantities: (1) the discretized excitation light field; and (2) the discretized Green's function of the emission light field. To further analyze the structure of the FDOT forward matrix, we define two new matrices. The first one is an $N_s \times N$ matrix with its i^{th} row being the excitation light field due to the i^{th} source:

$$\boldsymbol{\Phi} := \begin{bmatrix} \phi^1_{ex,1} & \cdots & \phi^1_{ex,N} \\ \vdots & \dots & \vdots \\ \phi^{N_s}_{ex,1} & \cdots & \phi^{N_s}_{ex,N} \end{bmatrix} \in \mathbb{R}^{N_s \times N}. \tag{6.1}$$

The second one is an $N_d \times N$ matrix with its j^{th} row being the Green's function of the emission light field corresponding to the j^{th} detector:

$$\boldsymbol{G} := \begin{bmatrix} g^1_{em,1} & \cdots & g^1_{em,N} \\ \vdots & \dots & \vdots \\ g^{N_d}_{em,1} & \cdots & g^{N_d}_{em,N} \end{bmatrix} \in \mathbb{R}^{N_d \times N}. \tag{6.2}$$

Let $\boldsymbol{a}_k$, $k = 1, ..., N$ be the k^{th} column of the FDOT forward matrix $\boldsymbol{A}$; and let

$$\begin{aligned} \boldsymbol{\phi}_k &= [\phi^1_{ex,k}, ..., \phi^{N_s}_{ex,k}]^T \\ \boldsymbol{g}_k &= [g^1_{em,k}, ..., g^{N_d}_{em,k}]^T \end{aligned} \tag{6.3}$$

be the k^{th} column of $\boldsymbol{\Phi}$ and $\boldsymbol{G}$, respectively. We observe $\boldsymbol{a}_k$ is the Kronecker product of $\boldsymbol{\phi}_k$ and $\boldsymbol{g}_k$, and write

$$\boldsymbol{a}_k = \boldsymbol{\phi}_k \otimes \boldsymbol{g}_k, \quad k = 1, ..., N. \tag{6.4}$$

Thus, $\boldsymbol{A}$ is the columnwise Knocker product of $\boldsymbol{\Phi}$ and $\boldsymbol{G}$:

$$\boldsymbol{A} = \boldsymbol{\Phi} * \boldsymbol{G} = [\boldsymbol{\phi}_1 \otimes \boldsymbol{g}_1, \boldsymbol{\phi}_2 \otimes \boldsymbol{g}_2, ..., \boldsymbol{\phi}_N \otimes \boldsymbol{g}_N] \tag{6.5}$$

where $*$ denotes the columnwise Kronecker product. As a result of (6.4), we have

$$\begin{aligned} \|\boldsymbol{a}_k\|_2^2 &= \boldsymbol{a}_k \cdot \overline{\boldsymbol{a}_k} = (\boldsymbol{\phi}_k \otimes \boldsymbol{g}_k) \cdot (\overline{\boldsymbol{\phi}_k \otimes \boldsymbol{g}_k}) \\ &= \|\boldsymbol{\phi}_k\|_2^2 \|\boldsymbol{g}_k\|_2^2. \end{aligned} \tag{6.6}$$

Furthermore,

$$\begin{aligned} \langle \boldsymbol{a}_p,\ \boldsymbol{a}_q \rangle &= \boldsymbol{a}_p \cdot \overline{\boldsymbol{a}_q} = (\boldsymbol{\phi}_p \otimes \boldsymbol{g}_p) \cdot (\overline{\boldsymbol{\phi}_q \otimes \boldsymbol{g}_q}) \\ &= (\boldsymbol{\phi}_p \cdot \overline{\boldsymbol{\phi}_q})(\boldsymbol{g}_p \cdot \overline{\boldsymbol{g}_q}) \\ &= \langle \boldsymbol{\phi}_p,\ \boldsymbol{\phi}_q \rangle \langle \boldsymbol{g}_p,\ \boldsymbol{g}_q \rangle. \end{aligned} \tag{6.7}$$

Let $r_{A_{p,q}}$, $r_{\phi_{p,q}}$ and $r_{g_{p,q}}$ be the normalized inner product of the p^{th} and q^{th} columns of $\boldsymbol{A}$, $\boldsymbol{\Phi}$ and $\boldsymbol{G}$, respectively. Combining (6.6) and (6.7), we obtain

$$r_{A_{p,q}} = r_{\phi_{p,q}} r_{g_{p,q}}. \tag{6.8}$$

Below we summarize two results on the coherence of the FDOT forward matrix.

Observation 1[2]*:*

The mutual coherence of $\boldsymbol{A}$ is upper bounded by the product of the mutual

---

[2] Note that since $\boldsymbol{\Phi}$ is composed of the excitation light field and $\boldsymbol{G}$ is composed of the discretized Green's function of the diffusion equation for the emission light field. Therefore, every entry in $\boldsymbol{\Phi}$ and $\boldsymbol{G}$ is non-negative. As a result, $r_{\phi_{p,q}} \geq 0$ and $r_{g_{p,q}} \geq 0$.

coherence of $\boldsymbol{G}$ and $\boldsymbol{\Phi}$.

$$\begin{aligned}\mathcal{M}(\boldsymbol{A}) &= \max_{p,q} |r_{A_{p,q}}| \\ &= \max_{p,q} |r_{\phi_{p,q}} r_{g_{p,q}}| = \max_{p,q} r_{\phi_{p,q}} r_{g_{p,q}} \\ &\leq (\max_{p,q} r_{\phi_{p,q}})(\max_{p,q} r_{g_{p,q}}) = \mathcal{M}(\boldsymbol{G})\mathcal{M}(\boldsymbol{\Phi}). \qquad (6.9)\end{aligned}$$

□.

Observation 2:

$$\begin{aligned}\mathcal{M}(\boldsymbol{A}) &\leq \mathcal{M}_1(k, \boldsymbol{A}) \\ &\leq K(N, N_s, N_d, k)(\|\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2 + \|\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}} - \boldsymbol{I}_N\|_F^2), \\ &\quad \text{for } k = 1, ..., N-1. \qquad (6.10)\end{aligned}$$

where $\|\boldsymbol{E}\|_F = \sqrt{\sum_{i,j} e_{i,j}^2}$ denotes the Frobenius norm of the matrix quantity $\boldsymbol{E}$; $K(N, N_s, N_d, k)$, $i = 1, 2$ is a constant that depends on N, N_s,N_d and k; $\tilde{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{G}}$ are $\boldsymbol{\Phi}$ and $\boldsymbol{G}$ matrices whose columns are normalized to unity, and $\boldsymbol{I}_N$ denotes the N-by-N identity matrix.

□.

See Appendix B for the derivation of (6.10), and an explicit expression for K.

We summarize the implications of the results and observations made in this section in the following list of remarks:
Remarks -

- The normalized inner product $r_{A_{p,q}}$ of the p^{th} and q^{th} columns of the FDOT forward matrix is the product $r_{\phi_{p,q}}$ and $r_{g_{p,q}}$, the normalized inner product of the p^{th} and q^{th} columns of $\boldsymbol{\Phi}$ and $\boldsymbol{G}$ matrices, respectively.

- To reduce the mutual coherence and cumulative coherence of the FDOT forward matrix $\boldsymbol{A}$, we wish to minimize $r_{A_{p,q}}$, which is equivalent to minimizing $r_{g_{p,q}}$ and $r_{\phi_{p,q}}$.

- The mutual coherence of the FDOT forward matrix is upper bounded by the product of the mutual coherence of $\boldsymbol{\Phi}$ and $\boldsymbol{G}$, i.e., $\mathcal{M}(\boldsymbol{A}) \leq \mathcal{M}(\boldsymbol{\Phi})\mathcal{M}(\boldsymbol{G})$.

- The sum of all the normalized inner products of the columns of the FDOT forward matrix is upper bounded by the sum of $\|\tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2$ and $\|\tilde{\boldsymbol{G}}^T\tilde{\boldsymbol{G}} - \boldsymbol{I}_N\|_F^2$. Thus, reducing $\|\tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2$ and $\|\tilde{\boldsymbol{G}}^T\tilde{\boldsymbol{G}} - \boldsymbol{I}_N\|_F^2$, reduces $r_{A_{p,q}}$, which in turn, reduces the cumulative coherence, $\mathcal{M}_1(k, \boldsymbol{A})$, and mutual coherence, $\mathcal{M}(\boldsymbol{A})$, of the FDOT forward matrix.

Without loss of generality for the rest of this chapter, we assume that the columns of $\boldsymbol{G}$ and $\boldsymbol{\Phi}$ are normalized to unity and drop tilde from $\tilde{\boldsymbol{G}}$ and $\tilde{\boldsymbol{\Phi}}$ to simplify our notation for rest of the chapter (see Appendix C for the equivalence of the underlying inverse problem).

6.3 The FDOT Forward Matrix Optimization

In this section, we focus on the optimization of the FDOT forward matrix using the results in the previous section. In particular, we minimize $\|\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \boldsymbol{I}_N\|_F$ and $\|\boldsymbol{G}^T\boldsymbol{G} - \boldsymbol{I}_N\|_F^2$ by designing two preconditioners on $\boldsymbol{\Phi}$ and $\boldsymbol{G}$ to reduce both the mutual coherence and average coherence of the FDOT forward matrix. Our preconditioner design follows a method similar to the one in [53].

The entries of the matrix $\boldsymbol{\Phi}$ represent the excitation field at each voxel location due to light sources. They are determined by the source function, the Green's function of (2.1) and (2.5) and the endogenous optical properties of the imaging domain. Thus, $\boldsymbol{\Phi}$ is a source related term. We refer to the preconditioner on $\boldsymbol{\Phi}$ as the *optical mask*. The optical mask is an $N_s \times N_s$ intensity matrix, with each row being a different intensity pattern applied to N_s point sources for N_s different illumination patterns.

The entries of the matrix $\boldsymbol{G}$ represent the emission field at the detectors due to a point source at each voxel location. They are determined by the Green's function of (2.2) and (2.6), and the endogenous optical properties of the imaging domain. Thus, $\boldsymbol{G}$ can be viewed as a measurement-related term. We refer to the preconditioner on $\boldsymbol{G}$ as the *measurement mask*. The measurement mask is an $N_d \times N_d$ linear filter applied to the measured data collected by N_d detectors corresponding to each illumination pattern.

In Subsection 6.3.1 we present the design of the optical mask; and in Subsection 6.3.2, we present the design of measurement mask. In Subsection 6.3.3, we describe the practical, algorithmic implementation of the FDOT imaging process with structured illumination and detection patterns.

6.3.1 Design of the Illumination Patterns - The Optical Mask

In diffuse optical tomography, the light sources are typically time multiplexed, i.e., operated one at a time, and the scattered light field due to each source is measured at the detectors. In designing the optical mask, we assume that all N_s sources transmit simultaneously into the imaging domain according to some intensity patterns designed to optimize the FDOT forward matrix. Furthermore, we assume that the number of illumination patterns is the same as the number of point sources. Thus, our objective is to determine $N_s \times N_s$ different light intensities to precondition $\mathbf{\Phi}$, and hence the FDOT forward matrix.

Let $\boldsymbol{M}_s$ denote the $N_s \times N_s$ optical mask. Let $\boldsymbol{m}_k$ denote each row of $\boldsymbol{M}_s$, i.e.,

$$\boldsymbol{M}_s = [\boldsymbol{m}_1{}^T, \boldsymbol{m}_2{}^T, \, ... \, \boldsymbol{m}_{N_s}{}^T]^T. \tag{6.11}$$

We refer to $\boldsymbol{m}_k$, $k = 1, ..., N_s$ as the k^{th} illumination pattern for reasons that will be made clear below.

Recall that the excitation light field due to the i^{th} point source is given by

$$\phi_x^i(\boldsymbol{r}) = \int_\Omega g_x(\boldsymbol{r}, \boldsymbol{r}')s^i(\boldsymbol{r}')d\boldsymbol{r}'. \tag{6.12}$$

Without loss of generality, we assume that the initial light sources, $s^i(\boldsymbol{r})$, are unit impulse functions,

$$s^i(\boldsymbol{r}) = \delta(\boldsymbol{r} - \boldsymbol{r}_i) \tag{6.13}$$

where $\boldsymbol{r}_i$ denotes the location of the i^{th} point source.

Let $m_{k,i}$ be the intensity of the i^{th} source, $i = 1, ..., N_s$, for the k^{th} illumination pattern, i.e.,

$$\boldsymbol{m}_k = [m_{k,1}, m_{k,2}, \, ... \, m_{k,N_s}]. \tag{6.14}$$

Then, if all point sources transmit simultaneously, the source illuminating the imaging domain can be expressed as

$$Q^k(\boldsymbol{r}) = \sum_{i=1}^{N_s} m_{k,i} s^i(\boldsymbol{r}). \tag{6.15}$$

Thus, the resulting excitation light field due to the k^{th} illumination pattern becomes

$$\begin{aligned} \phi_{pre}^k(\boldsymbol{r}) &= \int_\Omega g_x(\boldsymbol{r}, \boldsymbol{r}') Q^k(\boldsymbol{r}') d\boldsymbol{r}' \\ &= \sum_{i=1}^{N_s} m_{k,i} \int_\Omega g_x(\boldsymbol{r}, \boldsymbol{r}') s^i(\boldsymbol{r}') d\boldsymbol{r}' \\ &= \sum_{i=1}^{N_s} m_{k,i} \phi_x^i(\boldsymbol{r}). \end{aligned} \tag{6.16}$$

Reexpressing (6.16) in matrix notation introduced in (6.11) and (6.14), we write the new matrix composed of the excitation light field due to sources Q^k $(k = 1, ..., N_s)$ as follows:

$$\boldsymbol{\Phi}_{pre} = \boldsymbol{M}_s \boldsymbol{\Phi}, \tag{6.17}$$

where $\boldsymbol{\Phi}_{pre}$ is now the preconditioned $\boldsymbol{\Phi}$ matrix. From (6.17), clearly the optical mask $\boldsymbol{M}_s$ can be designed to precondition $\boldsymbol{\Phi}$. We design $\boldsymbol{M}_s$ to minimize the normalized inner products between different columns of $\boldsymbol{\Phi}_{pre}$. Or equivalently, we seek to find $\boldsymbol{M}_s$ such that the Gramm matrix ${\boldsymbol{\Phi}_{pre}}^T \boldsymbol{\Phi}_{pre}$ is as close to identity as possible,

$${\boldsymbol{\Phi}_{pre}}^T \boldsymbol{\Phi}_{pre} = \boldsymbol{\Phi}^T \boldsymbol{M}_s^T \boldsymbol{M}_s \boldsymbol{\Phi} \approx \boldsymbol{I}_N. \tag{6.18}$$

To find such a matrix $\boldsymbol{M}_s$, we consider the following optimization problem:

$$\min_{\boldsymbol{M}_s} \|(\boldsymbol{M}_s \boldsymbol{\Phi})^T \boldsymbol{M}_s \boldsymbol{\Phi} - \boldsymbol{I}_N\|_F. \tag{6.19}$$

Note that (6.19) attempts to minimize not only the greatest normalized inner product between two columns of $\boldsymbol{\Phi}_{pre}$, but all the normalized inner products between different columns of $\boldsymbol{\Phi}_{pre}$.

To solve (6.19), we multiply (6.18) by $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^T$ on both sides,

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{M}_s^T\boldsymbol{M}_s\boldsymbol{\Phi}\boldsymbol{\Phi}^T \approx \boldsymbol{\Phi}\boldsymbol{\Phi}^T, \tag{6.20}$$

and consider the singular value decomposition of $\boldsymbol{\Phi}$,

$$\boldsymbol{\Phi} = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{V}_s, \tag{6.21}$$

where $\boldsymbol{\Sigma}_s$ is the $N_s \times N_s$ diagonal matrix containing the singular values of $\boldsymbol{\Phi}$. Let $\boldsymbol{\Lambda}_s = \boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^T$. Substituting (6.21) into (6.20), and simplifying (6.20), we obtain

$$\boldsymbol{\Lambda}_s\boldsymbol{U}_s^T\boldsymbol{M}_s^T\boldsymbol{M}_s\boldsymbol{U}_s\boldsymbol{\Lambda}_s \approx \boldsymbol{\Lambda}_s. \tag{6.22}$$

Thus, we choose $\boldsymbol{M}_s$ as

$$\boldsymbol{M}_s = \boldsymbol{\Lambda}_s^{-1/2}\boldsymbol{U}_s^T. \tag{6.23}$$

The new $\boldsymbol{\Phi}$ matrix after the application of $\boldsymbol{M}_s$ becomes

$$\boldsymbol{\Phi}_{pre} = \boldsymbol{M}_s\boldsymbol{\Phi} = \boldsymbol{\Lambda}_s^{-1/2}\boldsymbol{U}_s^T\boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{V}_s = \boldsymbol{\Lambda}_s^{-1/2}\boldsymbol{\Sigma}_s\boldsymbol{V}_s. \tag{6.24}$$

The matrix $\boldsymbol{M}_s$ may have negative entries. However, since the intensity of each light source must be non-negative, we decompose $\boldsymbol{M}_s$ into two parts:

$$\boldsymbol{M}_s = \boldsymbol{M}_s^{(+)} - (-\boldsymbol{M}_s^{(-)}), \tag{6.25}$$

where $\boldsymbol{M}_s^{(+)}$ and $\boldsymbol{M}_s^{(-)}$ contain the non-negative and negative entries in $\boldsymbol{M}_s$, respectively. Instead of applying the illumination patterns given by $\boldsymbol{M}_s$, we sequentially apply the rows of $\boldsymbol{M}_s^{(+)}$ and $(-\boldsymbol{M}_s^{(-)})$, and take the differences of the measurements.

Let $m_{k,i}^{(\pm)}$ denote the $(k,i)^{th}$ entry of $\boldsymbol{M}_s^{(\pm)}$, which is the intensity of the i^{th} point source in the k^{th} illumination pattern. Thus, the light source generated by the k^{th} illumination pattern is given by

$$Q^k(\boldsymbol{r}) = \sum_{i=1}^{N_s}(m_{k,i}^{(+)} - (-m_{k,i}^{(-)}))s^i(\boldsymbol{r}), \tag{6.26}$$

and the resulting excitation light field due to the k^{th} illumination pattern becomes

$$\boldsymbol{\phi}_{pre}^{(k)}(\boldsymbol{r}) = \sum_{i=1}^{N_s} m_{k,i}^{(+)} \boldsymbol{\phi}_x^i(\boldsymbol{r}) - \sum_{i=1}^{N_s} (-m_{k,i}^{(-)}) \boldsymbol{\phi}_x^i(\boldsymbol{r}), \quad k = 1, ..., N \quad \boldsymbol{r} \in \Omega. \tag{6.27}$$

6.3.2 Design of Detection Patterns - The Measurement Mask

In this section, our objective is to design a measurement mask, or a linear filter, that will be applied to the measured data to reduce the normalized inner product between the columns of $\boldsymbol{G}$, i.e., $r_{g_{p,q}}$.

Recall from Chapter 2 that the measurements $\boldsymbol{\Gamma}^i$ due to the i^{th} light source is given as $\boldsymbol{\Gamma}^i = \boldsymbol{B}^i \boldsymbol{x}$ where

$$\boldsymbol{B}^i = [\boldsymbol{g}_1 \phi_{ex,1}^i, \boldsymbol{g}_2 \phi_{ex,2}^i, ..., \boldsymbol{g}_N \phi_{ex,N}^i].$$

If a linear transform is applied to the measurements, the new set of measurements becomes:

$$\boldsymbol{\Gamma}_{pre}^i = \boldsymbol{M}_d \boldsymbol{\Gamma}^i = \boldsymbol{M}_d \boldsymbol{B}^i \boldsymbol{x} \tag{6.28}$$

where $\boldsymbol{M}_d \in \mathbb{R}^{N_d \times N_d}$ is the linear (not necessarily Toeplitz) filter, which we refer to as the measurement mask. We note that

$$\boldsymbol{M}_d \boldsymbol{B}^i = [\boldsymbol{M}_d \boldsymbol{g}_1 \phi_{ex,1}^i, \boldsymbol{M}_d \boldsymbol{g}_2 \phi_{ex,2}^i, ..., \boldsymbol{M}_d \boldsymbol{g}_N \phi_{ex,N}^i]. \tag{6.29}$$

Let

$$\boldsymbol{G}_{pre} = \boldsymbol{M}_d \boldsymbol{G} \tag{6.30}$$

be the preconditioned $\boldsymbol{G}$ matrix. We choose to design $\boldsymbol{M}_d$ so that the normalized inner products between different columns of $\boldsymbol{G}$ are minimized.

Designing the measurement mask, $\boldsymbol{M}_d$, is similar to designing the optical mask. Following a step similar to the one in (6.20), we obtain

$$\boldsymbol{G}\boldsymbol{G}^T \boldsymbol{M}_d^T \boldsymbol{M}_d \boldsymbol{G}\boldsymbol{G}^T \approx \boldsymbol{G}\boldsymbol{G}^T. \tag{6.31}$$

Let $\boldsymbol{G} = \boldsymbol{U}_d \boldsymbol{\Sigma}_d \boldsymbol{V}_d^T$ be the singular value decomposition of $\boldsymbol{G}$. We choose

$$\boldsymbol{M}_d = \boldsymbol{\Lambda}_d^{-1/2} \boldsymbol{U}_d^T \tag{6.32}$$

where $\boldsymbol{\Lambda}_d = \boldsymbol{\Sigma}_d \boldsymbol{\Sigma}_d^T$. The preconditioned $\boldsymbol{G}$ matrix after the application of $\boldsymbol{M}_d$ becomes

$$\boldsymbol{G}_{pre} = \boldsymbol{M}_d \boldsymbol{\Phi} = \boldsymbol{\Lambda}_d^{-1/2} \boldsymbol{U}_d^T \boldsymbol{U}_d \boldsymbol{\Sigma}_d \boldsymbol{V}_d = \boldsymbol{\Lambda}_d^{-1/2} \boldsymbol{\Sigma}_d \boldsymbol{V}_d. \tag{6.33}$$

Note that we have designed an $N_d \times N_d$ measurement mask $\boldsymbol{M}_d$ by preconditioning $\boldsymbol{G}$. In Chapter 5, an alternative approach in designing a measurement mask is to directly precondition the forward sensing matrix $\boldsymbol{A}$. Recall from Chapter 5, the measurement mask becomes

$$\boldsymbol{M}_A = (\boldsymbol{\Sigma}_A \boldsymbol{\Sigma}_A^T)^{-1/2} \boldsymbol{U}_A^T \tag{6.34}$$

where $\boldsymbol{U}_A \boldsymbol{\Sigma}_A \boldsymbol{V}_A^T$ is the singular value decomposition of $\boldsymbol{A}$. Note that applying $\boldsymbol{M}_A$ on $\boldsymbol{A}$ results in the pseudo-inverse solution of the linear system $\boldsymbol{y} \approx \boldsymbol{A}\boldsymbol{x}$. While applying $\boldsymbol{M}_G$ and $\boldsymbol{M}_s$ on $\boldsymbol{G}$ and $\boldsymbol{\Phi}$, respectively, does not result in the the pseudo-inverse of $\boldsymbol{A}$. Unlike $\boldsymbol{G}$, $\boldsymbol{A}$ usually has a large number of singular values equal or close to 0 in practice. This can be intuitively understood as follows: The elements of $\boldsymbol{G}$ represents the emission light field at each detector due to a unit point source at each voxel location. The elements of $\boldsymbol{A}$, on the other hand, represents the emission light field at each detector due to a source at each voxel location whose intensity is proportional to the excitation light field generated by a unit light source at the boundary. Thus, if the fluorophore is not sufficiently excited at all locations, $\boldsymbol{A}$ may have singular values close to zero resulting in zero emission field at all detector locations due to a non-zero fluorophore concentration. On the other hand, $\boldsymbol{G}$ is less likely to have singular values close to zero, since all voxel locations are assumed to be excited uniformly. As a result, the measurement mask $\boldsymbol{M}_A$ is typically ill-conditioned with its condition number approaching to infinity. Therefore, applying $\boldsymbol{M}_A$ is likely to severely amplify the additive noise. $\boldsymbol{M}_d$, on the other hand, is well-conditioned with a much smaller condition number. Additionally, designing $\boldsymbol{M}_d$ is computationally more efficient than designing $\boldsymbol{M}_A$, since $\boldsymbol{M}_d$ is N_s^2 times smaller in dimension than $\boldsymbol{M}_A$.

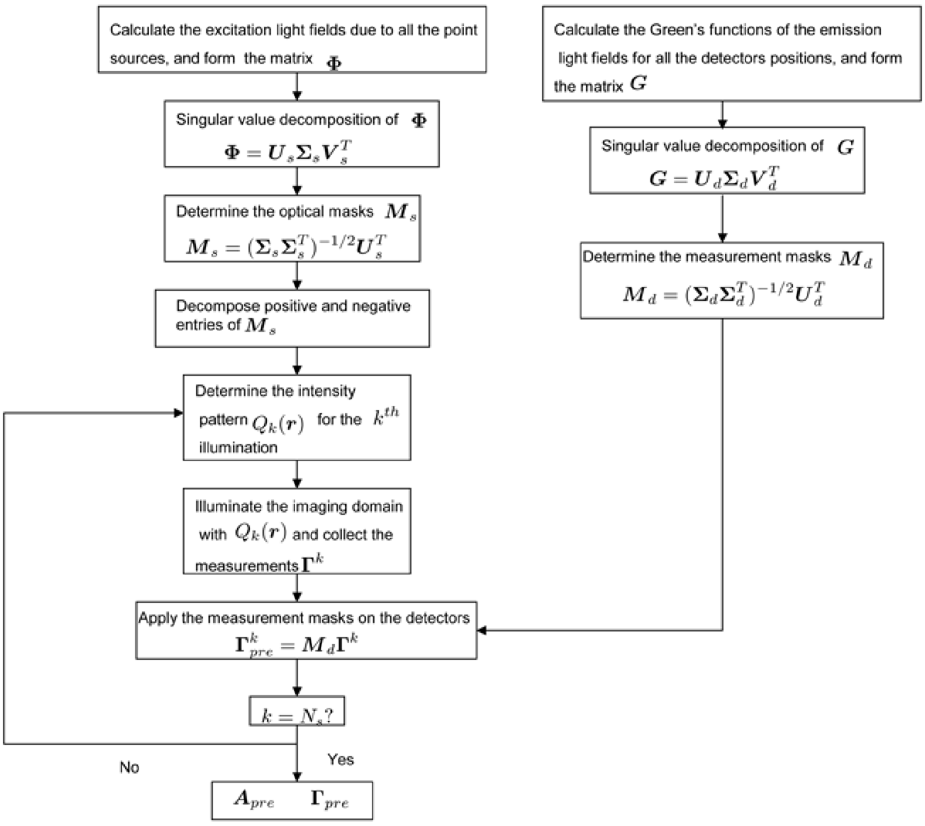


Figure 6.1: The block diagram of the FDOT imaging process with optical and measurement masks.

6.3.3 Forward Sensing Matrix Construction

Fig. 6.1 summarize Section 6.3.1 and 6.3.2 and shows the algorithmic description of the imaging process with optical and measurement masks. An overview of our method is demonstrated in Algorithm 3. The new forward sensing matrix $\boldsymbol{A}_{pre}$ composed of the columnwise Kronocker product of $\boldsymbol{\Phi}_{pre}$ and $\boldsymbol{G}_{pre}$,

$$\boldsymbol{A}_{pre} = [\boldsymbol{\phi}_{pre,1} \otimes \boldsymbol{g}_{pre,1}, ..., \boldsymbol{\phi}_{pre,N} \otimes \boldsymbol{g}_{pre,N}] \tag{6.35}$$

where $\boldsymbol{\phi}_{pre,i}$ and $\boldsymbol{g}_{pre,i}$, $i = 1, ..., N$ denote the columns of $\boldsymbol{\Phi}_{pre}$ and $\boldsymbol{G}_{pre}$.

Algorithm 3 Optical and measurement masks design and application.

```
STEP 1: Design the optical masks:
    · Calculate the excitation light fields, and form the matrix $\boldsymbol{\Phi}$;
    · Singular value decomposition of $\boldsymbol{\Phi}$: $\boldsymbol{\Phi} = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{V}_s^T$;
    · Calculate the optical maskS: $\boldsymbol{M}_s = (\boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^T)^{-1/2}\boldsymbol{U}_s^T$;
    · Decompose $\boldsymbol{M}_s$ into positive and negative parts: $\boldsymbol{M}_s = \boldsymbol{M}_s^{(+)} - (-\boldsymbol{M}_s^{(-)})$.
STEP 2: Design the measurement masks:
    · Calculate the Green's function, and form the matrix $\boldsymbol{G}$;
    · Singular value decomposition of $\boldsymbol{G}$: $\boldsymbol{G} = \boldsymbol{U}_d\boldsymbol{\Sigma}_d\boldsymbol{V}_d^T$;
    · Calculate the measurement masks: $\boldsymbol{M}_d = (\boldsymbol{\Sigma}_d\boldsymbol{\Sigma}_d^T)^{-1/2}\boldsymbol{U}_d^T$.
STEP 3: Application of the optical and measurement masks:
while $k \leq N_s$ do
  Get the $k$th row $\boldsymbol{m}_k^{(+)}$ ($\boldsymbol{m}_k^{(-)}$) from $\boldsymbol{M}_s^{(+)}$ ($-\boldsymbol{M}_s^{(-)}$);
  Set the intensity of point sources as the values in $\boldsymbol{m}_k^{(+)}$ ($\boldsymbol{m}_k^{(-)}$), and illuminate the imaging domain;
  Collect the measurements $\boldsymbol{\Gamma}^{k+}$ ($\boldsymbol{\Gamma}_{k-}$), $\boldsymbol{\Gamma}^k = \boldsymbol{\Gamma}^{k+} - \boldsymbol{\Gamma}^{k-}$;
  $\boldsymbol{\Gamma}_{pre}^k = \boldsymbol{M}_d\boldsymbol{\Gamma}^k$;
end while
STEP 4: Forming the forward sensing matrix and measurement vector:
    · $\boldsymbol{\Phi}_{pre} = \boldsymbol{M}_s\boldsymbol{\Phi}$, $\boldsymbol{G}_{pre} = \boldsymbol{M}_d\boldsymbol{G}$.
    · $\boldsymbol{A}_{pre} = [\boldsymbol{\phi}_{pre,1} \otimes \boldsymbol{g}_{pre,1}, ..., \boldsymbol{\phi}_{pre,N} \otimes \boldsymbol{g}_{pre,N}]$;
    · $\boldsymbol{\Gamma}_{pre} = [{\boldsymbol{\Gamma}_{pre}^{1}}^{T}, ... {\boldsymbol{\Gamma}_{pre}^{N_s}}^{T}]^T$.
```

6.4 Numerical Simulations

In this section, we extensive present numerical simulations to demonstrate the effects of the optical and measurement masks on the coherence of the forward sensing matrix as well as on the quality of the reconstructed FDOT images.

6.4.1 Simulation Setup

We set up a $6 \times 6 \times 6$ cm^3 cubic phantom shown in Fig.6.2. We set the background absorption coefficient to $\mu_{am} = 0.05$ cm^{-1}, and the diffusion coefficient to $D(\boldsymbol{r}) = 0.04$ cm^{-1} at both the excitation and emission wavelengths. The circular heterogeneity with radius r simulated the fluorophore concentration with fluorophore absorption coefficient $\mu_{axf} = 0.005$ cm^{-1}. 48 sources (shown in squares) and 48 detectors (shown in triangles) are uniformly placed at the top and bottom of the imaging domain. We discretized the imaging domain into $20 \times 20 \times 20$ voxels. Thus, the forward sensing matrix is 2304 by 8000. We simulated different data sets corresponding to 3 different values of the radius r : 0.5, 0.75, and 1.0 cm.

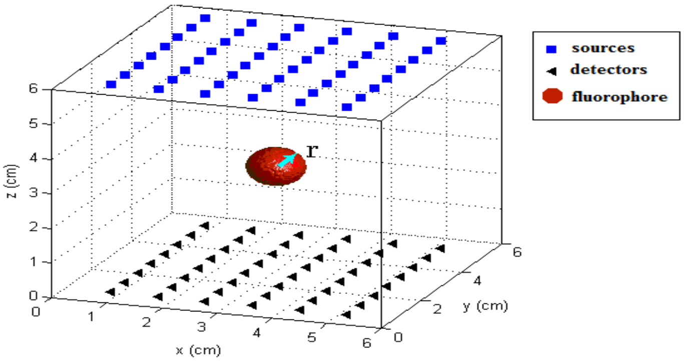
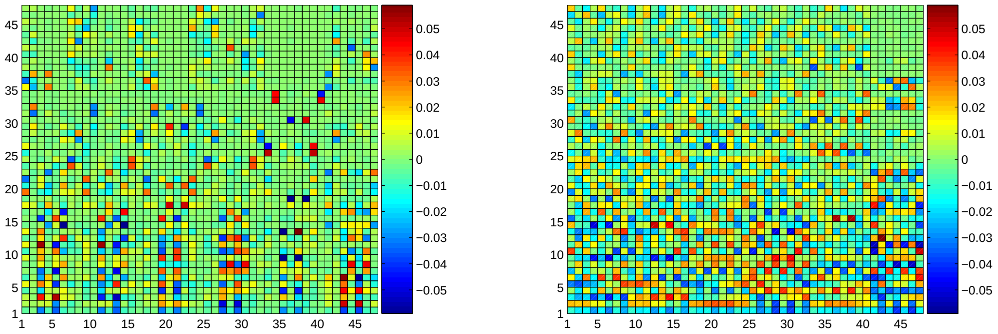


Figure 6.2: An illustration of the phantom and source-detector configuration.



(a) Optical mask

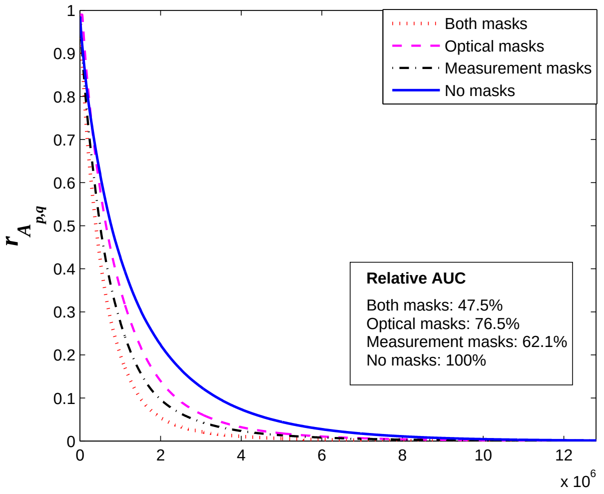
(b) Measurement mask

Figure 6.3: The optical and measurement masks used with simulated data.

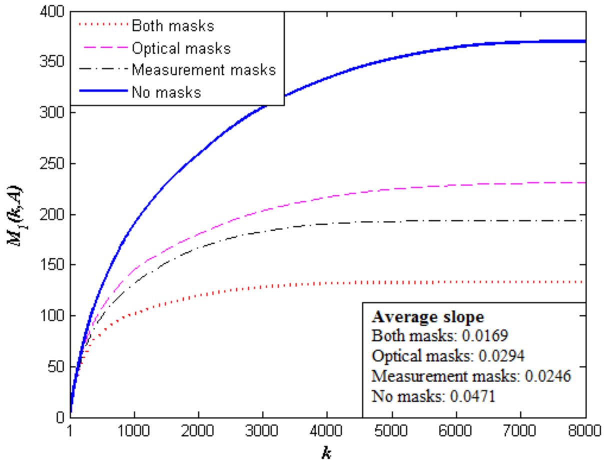
To simulate the measurement noise, we assume that, when a sufficiently large number of photons are detected, the noise at each detector can be approximated by a Gaussian random variable with its variance proportional to the magnitude of the detector reading. The SNR of the measurements is defined as

$$SNR = 20 \log_{10} \frac{\|\mathbf{\Gamma}\|_2}{\|\boldsymbol{\epsilon}\|_2}. \tag{6.36}$$

We simulated three sets of noise contaminated measurements with approximately 10%, 3% and 1% noise, corresponding to the SNR value of 20 dB, 30 dB and 40 dB, respectively.



(a) Top 40% values of all the normalized inner products



(b) Cumulative coherence

Figure 6.4: The normalized inner products and cumulative coherence of the forward matrix before and after applying optical and measurement masks.

6.4.2 Simulation Results - Coherence of the FDOT Forward Matrix

We evaluated the coherence of the forward matrix for four different cases: (1) no masks corresponding to the time-multiplexed point source configuration with unit intensity; (2) only optical mask $\boldsymbol{M}_s$; (3) only measurement mask $\boldsymbol{M}_d$; and (4) both optical and measurement masks. Fig. 6.3(a) and 6.3(b) show $\boldsymbol{M}_s$ and $\boldsymbol{M}_d$ matrices, respectively. Fig. 6.4(a) shows the largest 40% of all the normalized inner products between different columns of the forward sensing matrix, arranged in a descending order. The remaining 60% of the normalized inner products between different columns are close to 0. Clearly, application of either or both masks reduce the large correlations between different columns of $\boldsymbol{A}$. We observe that the measurement mask is more effective than the optical mask alone in reducing the coherence of the forward matrix. However, the application of both masks provides significantly better improvement than that of either masks alone. To quantify the improvement, we computed the *area under the curves* (AUC) for each case, and the relative percentages of AUC with respect to the no-mask case. The results of relative AUC are given in the box in Fig. 6.4(a). We see that the optical and measurement masks alone reduce the AUC to 76.5% and 62.1%, respectively as compared to the no-mask case. When both masks are applied, the AUC reduces to 47.5%.

Fig. 6.4(b) shows the plot of the cumulative coherence $\mathcal{M}_1(k, \boldsymbol{A})$ as a function of k. As stated before, the cumulative coherence is a nondecreasing function of k. When $\mathcal{M}_1(k, \boldsymbol{A})$ increases slowly, the forward matrix is said to be quasi-incoherent [77]. From Fig. 6.4(b), when optical or measurement masks are applied to the light sources and detectors, the cumulative coherence increases much slower as a function of k than that of no-mask case. The average slope of each curve is provided in the box in Fig. 6.4(b) for a quantitative comparison.

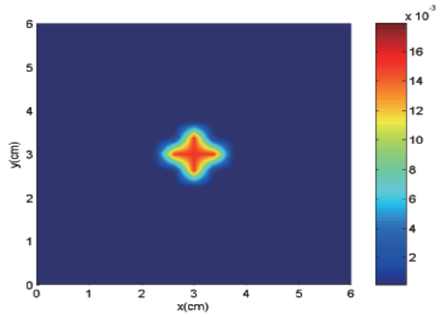


Figure 6.5: The cross section of the phantom at $z = 3$ (middle) ($r = 0.5$cm).

6.4.3 Simulation Results - Image Reconstruction

In this simulation, we reconstructed the fluorophore concentration map at 3 different sparsity levels corresponding to 3 different radii using measurements at 3 different noise levels. In each case, we used the forward matrix without applying any masks, with either optical or measurement masks alone, and with both masks applied simultaneously. We used six different sparsity promoting reconstruction methods available in the CS literature (See Chapter 4 for more details.). Specifically, we used StOMP [62], ROMP [63], CoSaMP [65], BP-interior [67], IST [69], and GPSR [71]. The first three are greedy type algorithms, and the last three are convex relaxation based algorithms.

We used NMSE to measure the accuracy of the reconstructed images,

$$\mathrm{NMSE} = \frac{\|\boldsymbol{x}_{\mathrm{true}} - \boldsymbol{x}_{\mathrm{recon}}\|_2}{\#\text{ of voxels}}. \tag{6.37}$$

We averaged the NMSE of the reconstructed images over 50 realizations of noise. The results are tabulated in Table 6.1. The results show that, in general, the NMSE of the reconstructed images increases as the sparsity level of the signal increases. By applying the optical or measurement masks, the NMSE can be reduced for all reconstructions. Table 6.1 shows a greater reduction in NMSE for greedy-type algorithms than that of convex relaxation techniques.

Fig. 6.5 shows the cross sections of the reconstructed images at the middle of the imaging domain when the radius of the heterogeneity is $r = 0.5$ cm. The cross sections of the reconstructed fluorophore yield maps using greedy type algorithms are shown in Figure6.6, and those using convex relaxation techniques are shown in Figure6.7. For each type of algorithm, the application of the optical and measurement masks results in reconstructed images that are in better agreement with the original fluorophore yield map. The visual improvements are most obvious for the greedy algorithms, consistent with the NMSE results. In the greedy type algorithms, the support of the signal is determined by selecting the columns of the forward matrix that have the greatest correlation with the measurements. The reduction of the normalized inner products between different columns in the forward matrix has a direct effect on the column selection procedure at each iteration.

Table 6.1: NMSE of the reconstructed images (10^{-4}).

Algorithms	$r = 0.5$ cm			$r = 0.75$ cm			$r = 1.0$ cm		
StOMP	40 dB	30 dB	20dB	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB
No masks	2.05	6.47	12.1	3.87	14.5	28.0	9.44	31.8	63.7
$\boldsymbol{M}_s$	0.98	2.05	7.04	1.90	6.47	13.9	5.00	13.2	33.3
$\boldsymbol{M}_d$	0.90	1.95	6.41	1.76	5.77	9.26	4.37	11.7	33.0
Both masks	0.55	1.31	2.58	0.93	2.63	5.55	2.12	7.67	16.4
ROMP	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB
No masks	1.66	5.85	13.2	3.68	13.0	31.6	8.01	29.0	64.2
$\boldsymbol{M}_s$	0.68	2.45	6.97	2.19	6.39	19.8	3.16	15.4	38.1
$\boldsymbol{M}_d$	0.54	2.54	6.62	1.89	5.19	16.9	3.71	13.9	35.3
Both masks	0.35	1.55	2.49	0.80	2.58	5.82	1.76	8.40	19.9
CoSaMP	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB
No masks	2.11	6.35	13.0	5.97	14.5	28.9	11.8	34.7	67.8
$\boldsymbol{M}_s$	1.69	2.11	5.98	3.69	8.96	13.4	4.29	16.1	35.6
$\boldsymbol{M}_d$	1.13	2.08	6.30	3.08	7.37	11.9	5.35	15.9	32.1
Both masks	0.61	1.37	3.45	1.31	2.77	6.08	2.66	8.06	18.5
BP-interior	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB
No masks	1.56	3.03	7.05	2.44	7.13	15.3	5.03	20.8	39.1
$\boldsymbol{M}_s$	1.18	2.31	4.85	1.97	5.07	10.2	4.02	10.8	23.4
$\boldsymbol{M}_d$	0.93	2.03	4.50	2.08	4.37	9.93	3.68	8.91	20.1
Both masks	0.39	1.28	1.98	1.09	2.06	3.85	1.84	5.18	9.11
IST	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB
No masks	1.20	3.55	7.49	3.51	6.97	19.2	6.38	23.3	49.9
$\boldsymbol{M}_s$	0.89	2.05	3.91	2.31	5.32	11.8	5.06	11.7	26.9
$\boldsymbol{M}_d$	0.92	1.79	3.61	1.93	4.19	10.0	4.41	9.95	24.3
Both masks	0.46	1.19	2.38	0.81	1.86	4.36	2.03	6.28	12.5
GPSR	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB	40 dB	30 dB	20 dB
No masks	1.29	3.28	7.07	3.41	7.13	17.1	5.51	22.2	46.6
$\boldsymbol{M}_s$	0.97	2.27	4.08	2.03	5.01	13.5	4.92	11.3	29.0
$\boldsymbol{M}_d$	0.82	1.76	3.26	1.50	3.76	10.0	4.83	8.74	24.8
Both masks	0.40	1.23	2.56	0.88	1.68	3.82	2.15	5.45	10.2

6.5 Silicon Phantom Experiment

In this section, we present the effect of using optical and measurement masks on the coherence of the forward sensing matrix and the reconstructed images using real data collected in a silicon phantom experiment.

6.5.1 Real Phantom Configuration

Figure6.8(a) shows an illustration of the cylindrical phantom used in the experiment. The phantom was made of silicone rubber with diameter of about 2 cm, and length of 4 cm. The phantom had homogeneous absorption coefficient $\mu_a = 0.2$ cm^{-1} and scattering coefficient $\mu_s' = 12\text{cm}^{-1}$ ($D = 1/3(\mu_a + \mu_s')$) at both the excitation and emission wavelengths (743 nm and 767 nm). The silicon phantom contained a hollow cylindrical tube in the middle with approximately 3mm in diameter. which was filled with intralipid and ink to mimic the same optical properties as the background. The intralipid and ink contained 1 micromolar of Cy7 as the fluorophore. The cross section of the fluorophore yield at $z = 1$ is shown in Figure 6.8(b).

The FDOT measurements were collected using the FDOT imaging system reported in [132], as illustrated in Figure6.9. Specifically, focused collimated laser beams were used as point light sources to excite the fluorophore. We had 60 point sources in total. The fluorescence measurements were collected by an electrically cooled CCD camera. The reading of the detector was recorded as the mean value of a subregion with 5×5 pixels around each detector location. We selected 60 detector locations. When the imaging domain is excited by an illumination pattern, i.e., multiple point sources of different intensities, we retrospectively approximated the measurements at the detectors by a combination of the measurements due to each point sources. We discretized the imaging domain into $20 \times 20 \times 20$ voxels. Thus, the forward sensing matrix is of dimension 3600 by 8000.

6.5.2 Coherence of the FDOT Forward Matrix

We evaluated the effect of the optical and measurement masks on the coherence of the forward matrix and on the quality of reconstructed images. Figure 6.10(a) and 6.10(b) show the optical and measurement masks, respectively. Figure 6.11(a) shows the largest 40% of all the normalized inner products between different columns of the forward sensing matrix, arranged in a descending order. Note that the remaining 60% of the normalized inner products between different columns are close to 0. Both the optical and measurement masks decrease the normalized inner products between different columns. The relative percentages of the AUC as compared to the no-masks case are given in the box in Figure 6.11(a). We see that the optical and measurement masks alone reduce the AUC to 74.4% and 64.1% as compared to the no-mask case. When both masks are applied, the AUC reduces to 45.0%.

Figure 6.11(b) shows the cumulative coherence $\mathcal{M}_1(k, \boldsymbol{A})$ of the forward matrix. We see a reduction in the average slope of each curve as shown in the box in Figure 6.11(b).

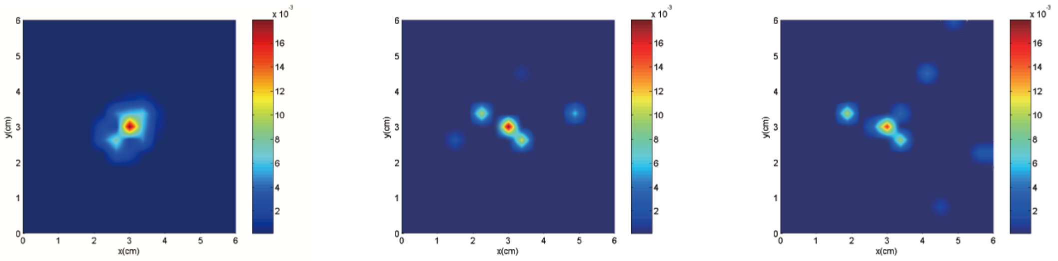
6.5.3 Image Reconstruction

The cross section of the original phantom at $z = 1$ (middle) is shown in Figure 6.8(b). The cross sections of the reconstructed fluorophore concentration maps of the silicon phantom using different greedy algorithms are shown in Figure6.12 and those using convex relaxation techniques are shown in Figure6.13. We observe that the application of optical and measurement masks improves the reconstruction results for all algorithms. Similar to the numerical simulation results, the improvements in the visual quality are most noticeable when the greedy algorithms are used.

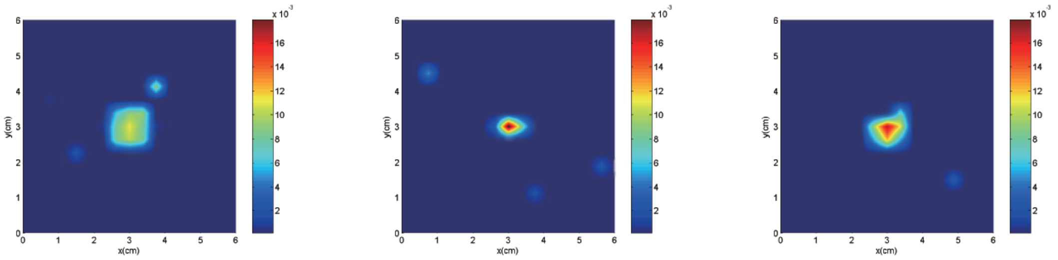
6.6 Conclusion

In this chapter, we presented a method of designing illumination and detection patterns for FDOT imaging to optimize the reconstruction of sparse fluorophore concentration maps. We showed that the FDOT forward matrix can be represented as the Kronecker product of two matrices. The first matrix is determined by the excitation light field, and the second matrix is determined by the Green's function of the emission light field. We showed that the incoherence of the FDOT forward matrix is related to the incoherence of these two matrices and design preconditioners to reduce the incoherence of these matrices. The preconditioners result in a set of intensity patterns for multiple spatially distributed sources illuminating the imaging domain simultaneously and a linear filter applied to the corresponding measurements collected at multiple spatially distributed detectors. To reconstruct the fluorophore concentration map, we used sparse signal recovery techniques, including both the greedy type and convex relaxation algorithms. We showed that the application of intensity patterns and filtering of the measurements in conjunction with sparse signal recovery techniques improves the visual quality of reconstructed images and reduces the mean square error in both numerical simulations and in a real phantom experiment.

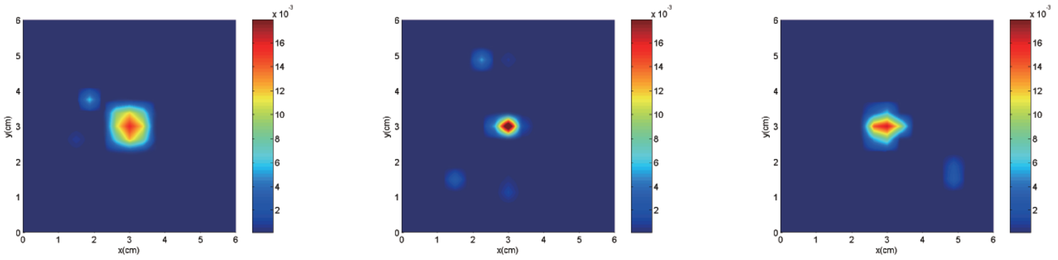
Our approach can be viewed as a method of determining the optimal source and detector locations and source intensities. Given a fluorophore designed to accumulate in certain regions and anatomical a priori information, optimal geometry for source-detector locations and source intensities can be determined via numerical simulations prior to the imaging process. This may eliminate unnecessary illumination, optimize imaging process and result in better image quality than the conventional FDOT imaging process.



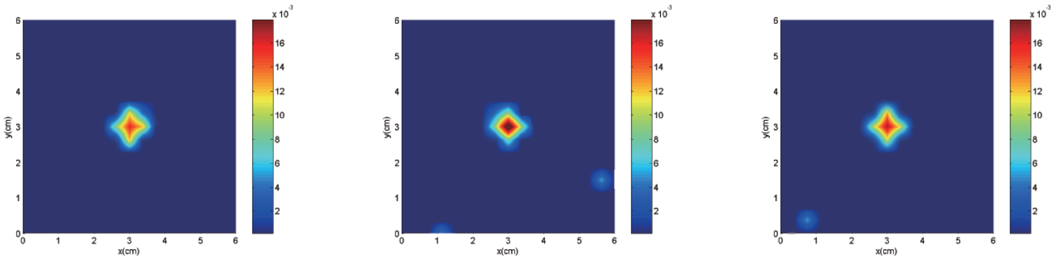
(a) StOMP (no masks) (b) ROMP (no masks) (c) CoSaMP (no masks)



(d) StOMP (optical masks) (e) ROMP (optical masks) (f) CoSaMP (optical masks)

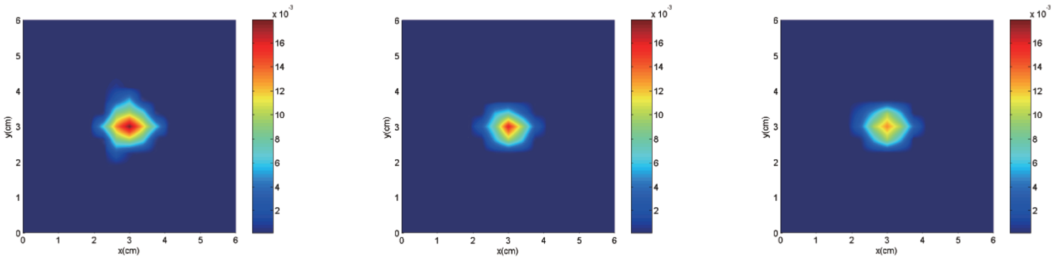


(g) StOMP (measurement mask) (h) ROMP (measurement mask) (i) CoSaMP (measurement mask)

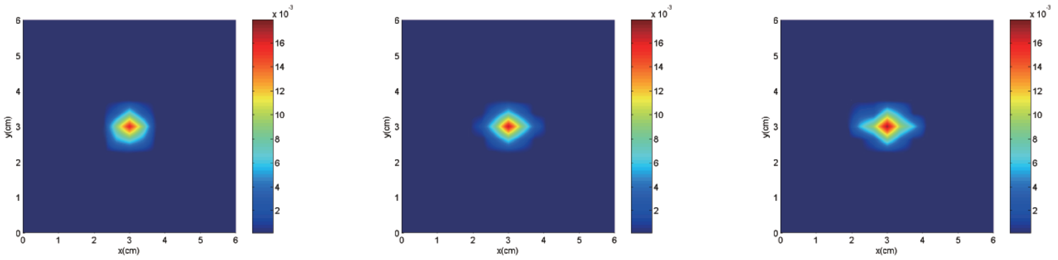


(j) StOMP (both masks) (k) ROMP (both masks) (l) CoSaMP (both masks)

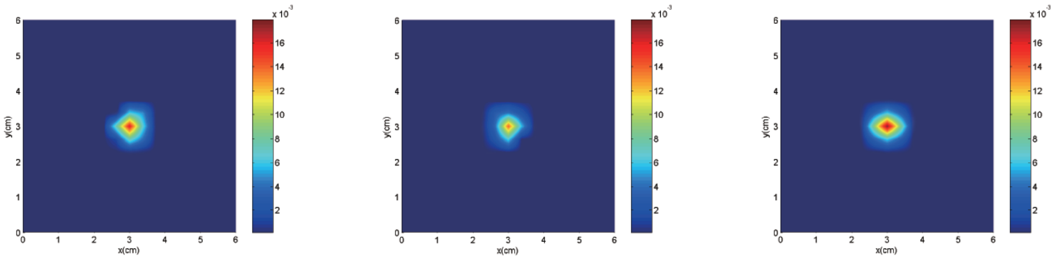
Figure 6.6: The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using greedy algorithms, $r = 0.5$ cm, 1% noise.



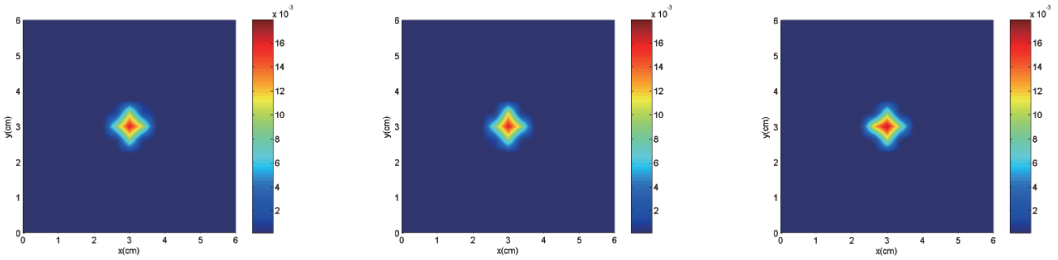
(a) BP-interior (no masks) (b) IST (no masks) (c) GPSR (no masks)



(d) BP-interior (optical masks) (e) IST (optical masks) (f) GPSR (optical masks)

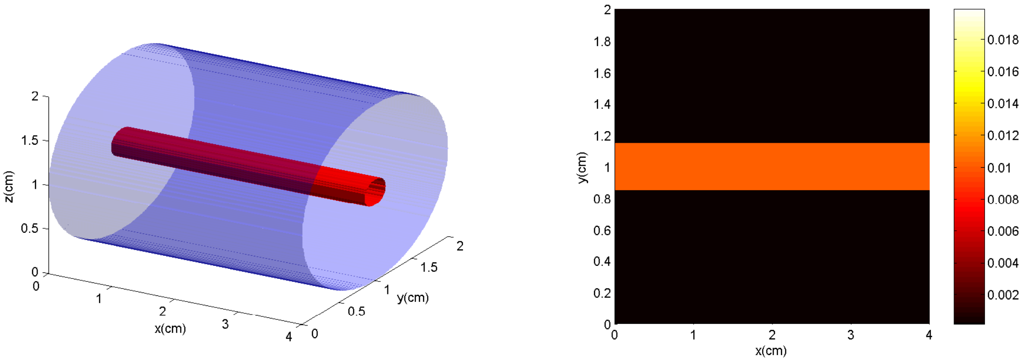


(g) BP-interior (measurement masks) (h) IST (measurement masks) (i) GPSR (measurement masks)



(j) BP-interior (both masks) (k) IST (both masks) (l) GPSR (both masks)

Figure 6.7: The cross sections of the reconstructed images at $z = 3$ (middle) of the phantom using convex relaxation algorithms, $r = 0.5$ cm, 1% noise.



(a) The silicon phantom

(b) The cross section of the fluorophore yield at $z = 1$ (middle) of the silicon phantom

Figure 6.8: The configuration of the real silicon phantom and the cross section of the fluorophore yield.

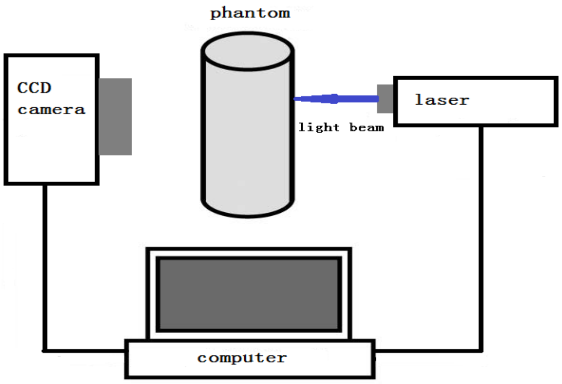
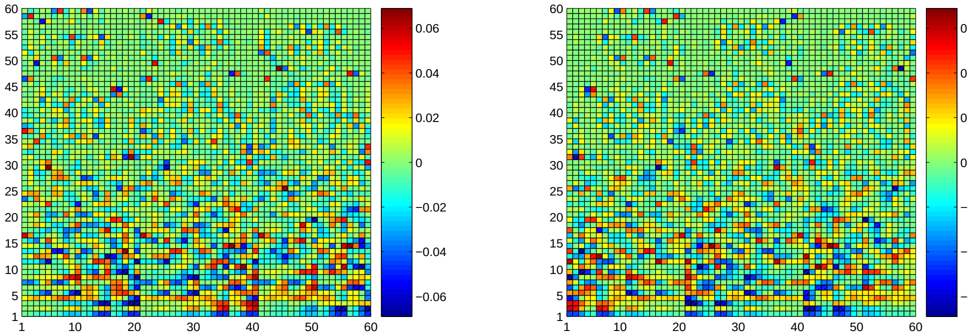


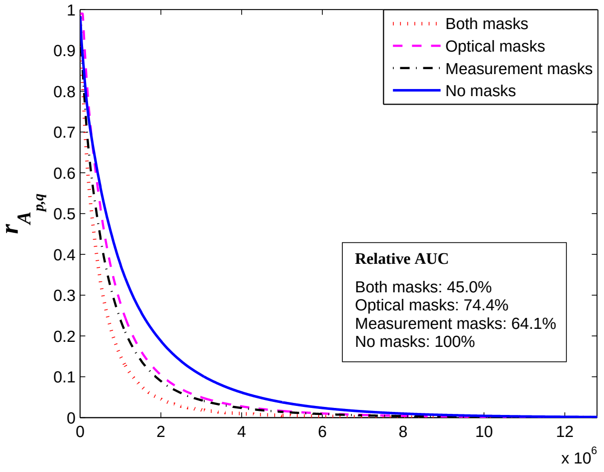
Figure 6.9: The imaging system used in the silicon phantom experiment.



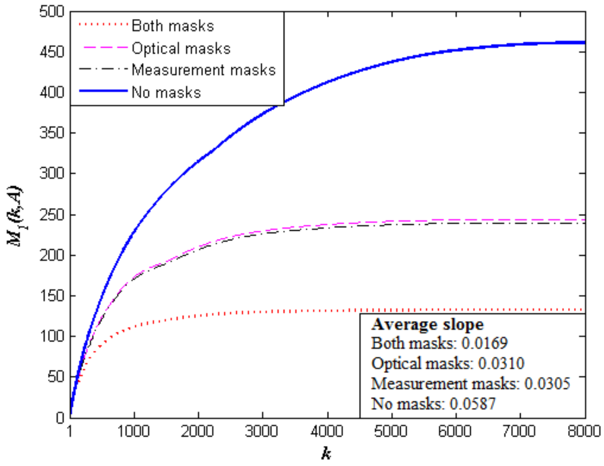
(a) Optical mask

(b) Measurement mask

Figure 6.10: The optical and measurement masks used with simulated data.

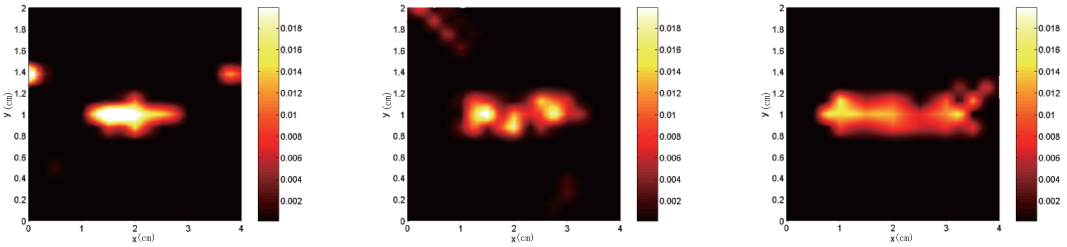


(a) Top 40% values of all the normalized inner products

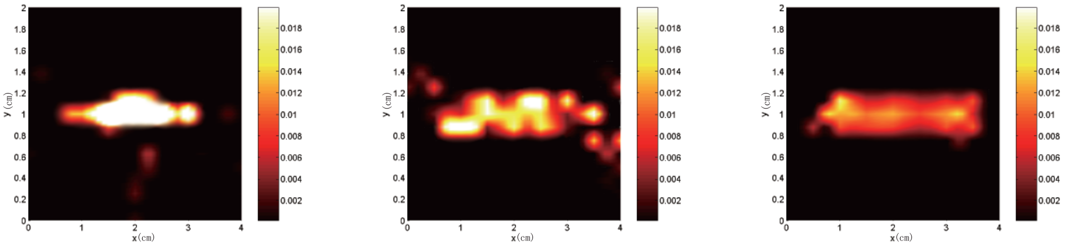


(b) Cumulative coherence

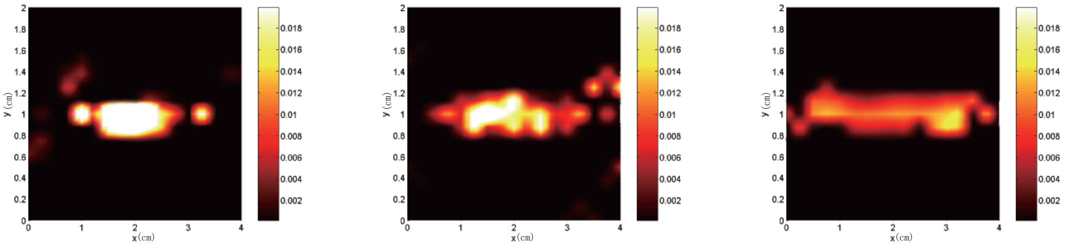
Figure 6.11: The normalized inner products and cumulative coherence of the forward matrix before and after applying optical and measurement masks.



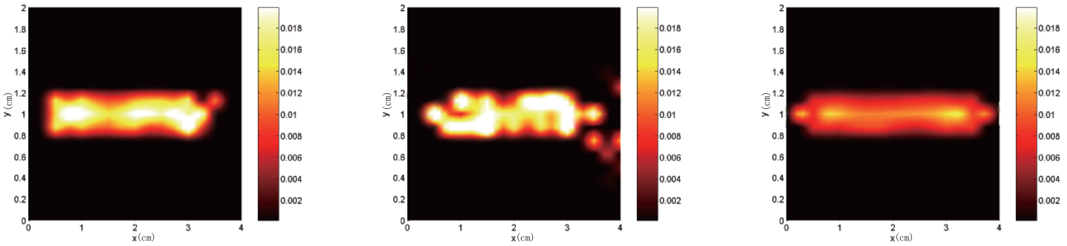
(a) StOMP (no masks) (b) ROMP (no masks) (c) CoSaMP (no masks)



(d) StOMP (optical masks) (e) ROMP (optical masks) (f) CoSaMP (optical masks)

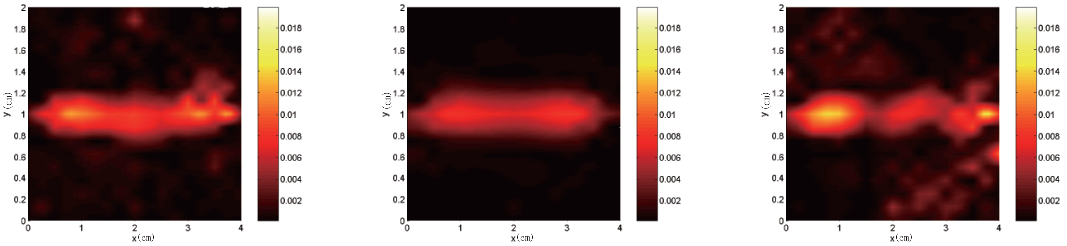


(g) StOMP (measurement masks) (h) ROMP (measurement masks) (i) CoSaMP (measurement masks)

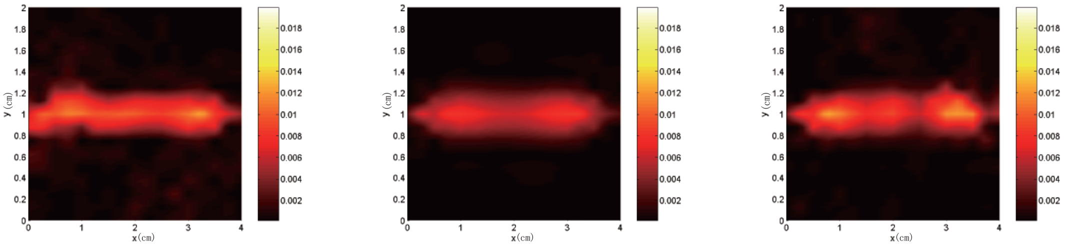


(j) StOMP (both masks) (k) ROMP (both masks) (l) CoSaMP (both masks)

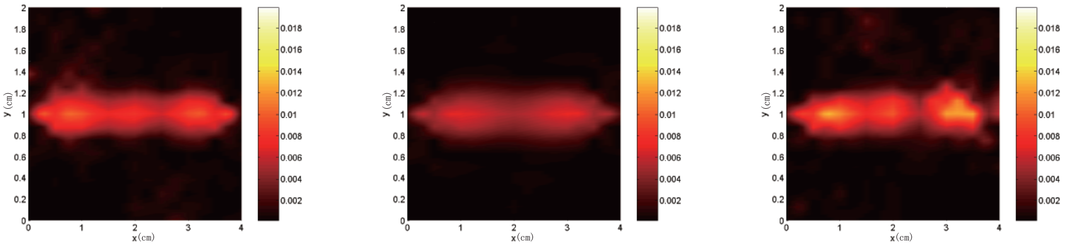
Figure 6.12: The cross sections of the reconstructed phantom at $z = 1$ (middle) using greedy algorithms.



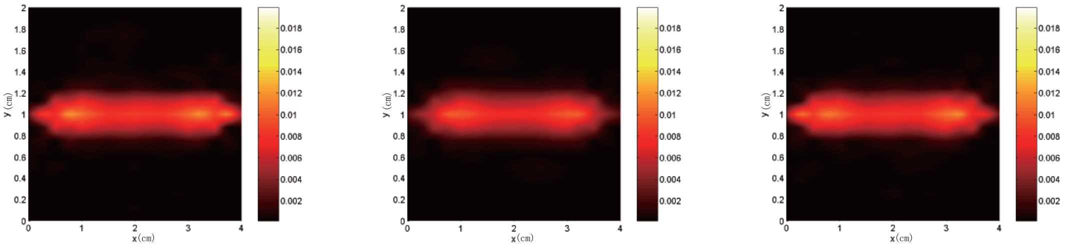
(a) BP-interior (no masks) (b) IST (no masks) (c) GPSR (no masks)



(d) BP-interior (optical masks) (e) IST (optical masks) (f) GPSR (optical masks)



(g) BP-interior (measurement masks) (h) IST (measurement masks) (i) GPSR (measurement masks)



(j) BP-interior (both masks) (k) IST (both masks) (l) GPSR (both masks)

Figure 6.13: The cross sections of the phantom at $z = 1$ (middle) using convex relaxation techniques.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this thesis, we addressed the ill-posedness of the FDOT inverse problem by (1) applying traditional regularization methods that suits to the specific nature of the fluorophore map reconstruction problem, and (2) exploiting the sparse nature of the fluorophore yield in the imaging domain based on CS framework. We modeled the light propagation in the imaging domain by two coupled diffusion equations in the frequency domain at the excitation and emission wavelength respectively. Based on the assumption of weak fluorophore, we formulated the FDOT image reconstruction as a linearized inverse problem, and used FEM to discritize the imaging domain.

In Chapter 3, we have reviewed a number of state-of-art regularization techniques in literature, and proposed a combined l_1-l_2-norm regularization technique. The proposed algorithm is capable of preserving the foreground region while smoothing out noise in the background region, which has improved upon state-of-the-art approaches.

In Chapter 4-6, we exploited the sparsity of the fluorophore concentration map and applied the CS framework for FDOT reconstruction. We noticed that the fluorophore distribution in the imaging domain is often very sparse. This is because that most optical fluorophore are "targeted probes" that are designed to accumulate in relatively small, specific regions in tissue. To summarize the results, we have covered the following topics:

1. In Chapter 4, we first studied various sparse signal recovery techniques in CS literature, and applied these techniques to reconstruct the fluorophore yield in the FDOT inverse problem. The results from both numerical simulation and real phantom data have shown that in general, the sparse signal recovery algorithms provide better reconstruction results compared to the traditional non-sparsity promoting regularization techniques. The relaxation algorithms

usually have better performances compared to the greedy-type algorithms.

2. Next, we addressed the incoherence property of the FDOT forward sensing matrix based on CS framework, and improved the incoherence of the FDOT forward sensing matrix for better FDOT reconstruction performance. In chapter 5, we designed a preconditioning matrix to apply on the FDOT forward sensing matrix. We have shown that the application of the preconditioning matrix reduces the coherence of the forward sensing matrix. The fluorophore concentration maps recovered from the preconditioned FDOT imaging system is more accurate compared to those recovered from the un-preconditioned imaging system.

3. In Chapter 6, we improved the incoherence of the FDOT forward sensing matrix by designing the light illumination and detection patterns. We decomposed the FDOT forward matrix into the Kronecker product of two underlying matrices. The first matrix depends on the excitation light fields and the configuration of light sources, and the second matrix depends on the emission light fields and the configuration of detectors. The incoherence of the FDOT forward sensing matrix is related to the incoherence of these two underlying matrices. By designing the light illumination and detection patterns, we applied two preconditioners to reduce the coherence of these two matrices, which, in turn, reduces the coherence of the FDOT forward matrix. We showed the improvement of the FDOT reconstruction performances using the illumination and detection patterns in 3D numerical simulation and real phantom experiments.

   Different from Chapter 5 where the FDOT forward sensing matrix is preconditioned directly, the method we used in Chapter 6 decomposes the forward matrix into two matrices with much smaller dimensions and condition numbers compared to the original one. This design is able to avoid noise amplification due to large condition number of the preconditioning matrix, thus is more robust to measurement noise.

7.2 Future Work

Based on the perspectives, methods, and algorithms developed in this thesis, we propose some future work in the following directions:

1. In this thesis, we focused on the linearized FDOT inverse problem. We note that the approach introduced in this thesis can be easily extended to nonlinear FDOT inverse problem within an iterative perturbation approach.

2. In FDOT reconstruction, we focused on estimating the fluorophore concentration using the zero frequency continuous wave method. However, the approach we presented in this thesis can be generalized to multiple frequency FDOT imaging system. The frequency information is related to the forward sensing matrix configuration, and can be utilized to improve the incoherence of the FDOT forward sensing matrix.

3. The discretization method has a directly impact on the FDOT forward sensing matrix formulation. In this thesis, we used finite element method with uniform mesh for discretization. An alternative approach is to use the adaptive mesh techniques. With adaptive mesh techniques, it is possible to eliminate coherent columns in the forward matrix which correspond to adjacent locations in the imaging domain, and thus improve the sparse fluorophore concentration map reconstruction.

4. In this thesis, we designed light illumination and detection patterns based on a number of point sources and detectors with pre-determined locations. However, the locations of light sources and detectors are also important factor in formulating the forward sensing matrix, which can be modeled as random variables. The incoherence of the forward sensing matrix can be further improved by optimization over these locations.

5. The fluorophore distribution map is typically sparse in nature, however, it can be further sparsified by a number of basis functions, e.g., the first order derivative basis functions, or the shape based basis functions. The sparse

signal recovery techniques developed in this thesis can be extended to combine various basis functions for sparse FDOT image reconstruction.

LITERATURE CITED

[1] V. Ntziachristos, A. Hielscher, A. Yodh, and B. Chance, “Diffuse optical tomography of highly heterogeneous media,” *IEEE Trans. Med. Imaging*, vol. 20, pp. 470–478, Jun. 2001.

[2] S. Arridge, “Optical tomography in medical imaging,” *Inverse Probl.*, vol. 15, no. 2, pp. 41–93, 1999.

[3] D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang, “Imaging the body with diffuse optical tomography,” *Signal Process. Mag.*, vol. 18, no. 6, pp. 57–75, Nov. 2001.

[4] B. Chance, “Optical method,” *Annu. Rev. Biophys.*, vol. 20, pp. 1–28, 1991.

[5] E. Sevick-Muraca, G. Lopez, T. Troy, J. Reynolds, and C. Hutchinson, “Fluorescence and absorption contrast mechanisms for biomedical optical imaging using frequency-domain techniques,” *Photochem. Photobiol.*, vol. 66, no. 1, pp. 55–64, 1997.

[6] J. Frangioni, “In vivo near-infrared fluorescence imaging,” *Curr. Opin. Chem. Biol.*, vol. 7, no. 5, pp. 626–634, Oct. 2003.

[7] E. Sevick-Muraca, J. Houston, and M. Gurfinkel, “Fluorescence-enhanced, near infrared diagnostic imaging with contrast agents,” *Curr. Opin. Chem. Biol.*, vol. 6, no. 5, pp. 642–650, Oct. 2002.

[8] R. Gillies, “In vivo molecular imaging,” *J. Cell. Biochem.*, vol. 87, pp. 231–238, 2002.

[9] C. Bremer, V. Ntziachristos, and R. Weissleder, “Optical-based molecular imaging: contrast agents and potential medical applications,” *Eur. Radiol.*, vol. 13, pp. 231–243, 2003.

[10] W. Chan, D. Maxwell, X. Gao, R. Bailey, M. Han, and S. Nie, “Luminescent quantum dots for multiplexed biological detection and imaging,” *Opin. Biotechnol.*, vol. 13, pp. 40–46, 2002.

[11] R. Tsien, “Building and breeding molecules to spy on cells and tumors,” *FEBS Lett.*, vol. 579, no. 4, pp. 927–932, 2005.

[12] M. Funovics, R. Weissleder, and C. Tung, “Protease sensors for bioimaging,” *Anal. Bioanal. Chem.*, vol. 377, no. 6, pp. 956–963, 2003.

[13] J. Reynolds, T. Troy, R. Mayer, A. Thompson, D. Waters, K. Cornell, P. Snyder, and E. Sevick-Muraca, “Imaging of spontaneous canine mammary tumors using uorescent contrast agents,” *Photochem. Photobiol.*, vol. 70, pp. 87–94, 1999.

[14] U. Mahmood, C. Tung, J. Bogdanov, and R. Weissleder, “Near-infrared optical imaging of protease activity for tumor detection,” *Radiology*, vol. 213, no. 3, pp. 866–870, 1999.

[15] A. Becker, C. Hessenius, K. Licha, B. Ebert, U. Sukowski, W. Semmler, B. Wiedenmann, and C. Grotzinger, “Receptor-targeted optical imaging of tumors with near-infrared uorescent ligands,” *Nat. Biotech.*, vol. 19, no. 3, pp. 327–331, 2001.

[16] J. Bugaj, S. Achilefu, R. Dorshow, and R. Rajagopalan, “Novel fluorescent contrast agents for optical imaging of in vivo tumors based on a receptortargeted dye-peptide conjugate platform,” *J. Biomed. Opt.*, vol. 6, no. 2, pp. 122–133, 2001.

[17] A. Corlu, R. Choe, T. Durduran, M. Rosen, M. Schweiger, S. Arridge, M. Schnall, and A. Yodh, “Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans,” *Opt. Express*, vol. 15, pp. 6696–6716, 2007.

[18] A. Pelegrin, S. Folli, F. Buchegger, J. Mach, G. Wagnieres, and H. van den

Bergh, “Antibody fluorescein conjugates for photoimmunodiagnosis of human colon carcinoma in nude mice,” *Cancer*, vol. 67, no. 10, pp. 2529–2537, 1991.

[19] B. Ballou, G. Fisher, T. Hakala, and D. Farkas, “Tumor detection and visualization using cyanine uorochrome-labeled antibodies,” *Biotechnol. Prog.*, vol. 13, no. 5, pp. 649–658, 1997.

[20] V. Ntziachristos, A. Yodh, M. Schnall, and B. Chance, “Concurrent mri and diffuse optical tomography of breast after indocyanine green enhancement,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 6, pp. 2767–2772, 2000.

[21] N. Unno, M. Suzuki, N. Yamamoto, K. Inuzuka, D. Sagara, M. Nishiyama, H. Tanaka, and H. Konno, “Indocyanine green fluorescence angiography for intraoperative assessment of blood flow: a feasibility study,” *Eur J Vasc Endovasc Surg.*, vol. 35, no. 2, pp. 205–207, 2008.

[22] V. Ntziachristos, C. Tung, C. Bremer, and R. Weissleder, “Fluorescence molecular tomography resolves protease activity in vivo,” *Nat. Med.*, vol. 8, no. 7, pp. 757 – 761, 2002.

[23] M. Patterson, B. Chance, and B. Wilson, “Time-resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties,” *Appl. Opt.*, vol. 28, no. 12, pp. 2331–2336, 1989.

[24] J. Wu, F. Partovi, M. Feld, and R. Rava, “Diffuse reflectance from turbid media: An analytic model of photon migration,” *Appl. Opt.*, vol. 32, no. 7, pp. 1115–1121, 1993.

[25] D. Boas, M. OLeary, B. Chance, and A. Yodh, “Scattering of diffuse photon density waves by spherical inhomogeneities within turbid media: Analytic solution and applications,” *in Proc. National Academy of Science*, vol. 91, no. 11, pp. 4887–4891, 1994.

[26] F. Martelli, A. Sassaroli, Y. Yamada, and G. Zaccanti, “Analytical approximate solutions of the time-domain diffusion equation in layered slabs,” *J. Opt. Soc. Am.*, vol. 19, no. 1, pp. 71–80, 2002.

[27] J. Ripoll, V. Ntziachristos, R. Carminati, and M. Nieto-Vesperinas, "Kirchhoff approximation for diffusive waves," *Phys Rev E Stat Nonlin Soft Matter Phys.*, vol. 64, no. 5 Pt 1, p. 051917, 2001.

[28] D. Boas, J. Culver, J. Stott, and A. Dunn, "Three dimensional monte carlo code for photon migration through complex heterogeneous media including the adult human head," *Opt. Express*, vol. 10, no. 3, pp. 159–69, 2002.

[29] V. Chernomordik, D. Hattery, D. Grosenick, H. Wabnitz, H. Rinneberg, K. Moesta, P. Schlag, and A. Gandjbakhche, "Quantification of optical properties of a breast tumour using random walk theory," *J. Biomed. Opt.*, vol. 7, no. 1, pp. 80–87, 2002.

[30] B. Brooksby, H. Dehghani, B. Pogue, and K. Paulsen, "Near infrared (nir) tomography breast image reconstruction with a priori structural information from mri: algorithm development for reconstructing heterogeneities," *IEEE J. Quantum Electron*, vol. 9, no. 2, pp. 199–209, Apr. 2003.

[31] A. Gibson, J. Riley, M. Schweiger, J. Hebden, S. Arridge, and D. Delpy, "A method for generating patient-specific finite element meshes for head modelling," *Phys. Med. Biol.*, vol. 48, no. 4, pp. 481–495, 2003.

[32] H. Dehghani, M. Doyley, B. Pogue, S. Jiang, J. Geng, and K. Paulsen, "Breast deformation modelling for image reconstruction in near infrared optical tomography," *Phys. Med. Biol.*, vol. 49, no. 7, pp. 1131–1145, 2004.

[33] L. Zhou and B. Yazici, "Discretization error analysis and adaptive meshing algorithms for fluorescence diffuse optical tomography in the presence of measurement noise," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1094–1111, Apr. 2011.

[34] M. Guven, B. Yazici, K. Kwon, E. Giladi, and X. Intes, "Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: Part i," *Inverse Probl.*, vol. 23, pp. 1115–1133, 2007.

[35] M. Guven, L. Zhou, L. Reilly-Raska, and B. Yazici, “Discretization error analysis and adaptive meshing algorithms for fluorescence diffuse optical tomography:part i,” *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 217–229, Feb. 2010.

[36] ——, “Discretization error analysis and adaptive meshing algorithms for fluorescence diffuse optical tomography:part ii,” *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 230–245, Feb. 2010.

[37] Y. Pei, H. Graberb, and R. Barbourb, “A fast reconstruction algorithm for implementation of time-series dc optical tomography,” in *Proc. SPIE. Optical Tomography Spectroscopy of Tissue*, vol. 4955, San Jose, CA, Jan. 2003, pp. 236–245.

[38] W. Zhu, Y. Wang, Y. Yao, J. Chang, H. Graber, and R. Barbour, “Iterative total least-squares image reconstruction algorithm for optical tomography by the conjugate gradient method,” *J. Opt. Soc. Am. A*, vol. 14, no. 4, pp. 799–807, 1997.

[39] M. O’Leary, D. Boas, B. Chance, and A. Yodh, “Experimental images of heterogenous turbid media by frequency-domain diffusing-photon tomography,” *Opt. Lett.*, vol. 20, no. 5, pp. 426–428, 1995.

[40] Y. Yao, Y. Wang, Y. Pei, W. Zhu, and R. Barbour, “Frequency domain optical imaging of absorption and scattering distributions by a born iterative method,” *J. Opt. Soc. Am. A*, vol. 14, no. 1, pp. 325–342, 1997.

[41] B. Pogue, T. McBride, J. Prewitt, U. Osterberg, and K. Paulsen, “Patially variant regularization improves diffuse optical tomography,” *Appl. Opt.*, vol. 38, no. 4, pp. 2950–2961, 1999.

[42] K. Paulsen and H. Jiang, “Enhanced frequency-domain optical image reconstruction in tissues through totalvariation minimization,” *Appl. Opt.*, vol. 35, no. 4, pp. 3447–3458, 1996.

[43] N. Cao, A. Nehorail, and M. Jacob, “Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm,” *Opt. express*, vol. 15, no. 21, pp. 13 695–13 708, 2007.

[44] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, 1st ed. New York, NY: Springer-Verlag, 2004.

[45] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[46] E. Candes and J. Romberg, “Quantitative robust uncertainty principles and optimally sparse decompositions,” *Foundations of Comput. Math.*, vol. 6, no. 2, pp. 227–254, 2006.

[47] S. Aeron, M. Zhao, and V. Saligrama, “On sensing capacity of sensor networks for the class of linear observation, fixed snr models,” in *Proc. IEEE Statistical Signal Process 14th Workshop*, Madison, WI, Aug. 2007, pp. 388–392.

[48] P. Mohajerani, A. Eftekhar, J. Huang, and A. Adibi, “Optimal sparse solution for fluorescent diffuse optical tomography: theory and phantom experimental results,” *Appl. Opt.*, vol. 46, no. 10, pp. 1679–1685, 2007.

[49] M. Mishali and Y. Eldar, “Reduce and boost: Recovering arbitrary sets of jointly sparse vectors,” *IEEE Trans. on Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Feb. 2008.

[50] V. Cevher, M. Duarte, and R. Baraniuk, “Distributed target localization via spatial sparsity,” in *Proc. European Signal Process. Conf. (EUSIPCO)*, Laussane, Switzerland, Aug. 2008, pp. 29–42.

[51] D. Donoho and M. Elad, “Optimally sparse representation in general (non-orthogonal) dictionaries via l_1 minimization,” *Proc. Nat. Aca. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[52] M. Elad, “Optimized projections for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[53] J. Duarte-Carvajalino and G. Sapiro, “Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization,” *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[54] W. Bajwa, R. Calderbank, and S. Jafarpour, “Why gabor frames? two fundamental measures of coherence and their role in model selection,” *J. Commun. Netw.*, vol. 12, no. 4, pp. 289–307, Aug. 2010.

[55] D. Donoho, M. Elad, and V. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[56] R. Roy, A. Godavarty, and E. Sevick-Muraca, “Fluorescence-enhanced optical tomography of a large tissue phantom using point illumination geometries,” *J. Biomed. Opt.*, vol. 11, no. 4, p. 044007, 2006.

[57] E. Hull, M. Nichols, and T. Foster, “Localization of luminescent inhomogeneities in turbid media with spatially resolved measurements of cw diffuse luminescence emittance,” *Appl. Opt.*, vol. 37, no. 13, pp. 2755–2765, 1998.

[58] J. Wu, Y. Wang, L. Perelman, I. Itzkan, I. Ramachandra, R. Dasari, and M. Feld, “Time-resolved multichannel imaging of fluorescent objects embedded in turbid media,” *Opt. Lett.*, vol. 20, no. 5, pp. 489–491, 1995.

[59] A. Milstein, “Imaging of near-infrared fluorescence, absorption, and scattering in turbid media,” Ph.D. dissertation, Purdue Univ., West Lafayette, Ind., Dec. 2004.

[60] J. Tropp, “Average-case analysis of greedy pursuit,” in *Proc. SPIE Wavelets XI*, San Diego, CA, Aug. 2005, pp. 340–350.

[61] T. Blumensath and M. Davies, “Gradient pursuits,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370–2382, Jun. 2008.

[62] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[63] D. Needell and R. Vershynin, “Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit,” *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Apr. 2009.

[64] ——, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[65] D. Needell and J. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[66] M. Elad, *Sparse and Redundant Representations: from Theory to Applications in Signal and Image Processing*, 1st ed. New York, NY: Springer, 2010.

[67] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Scientific Computing*, vol. 20, pp. 33–61, 1999.

[68] M. Figueiredo and R. Nowak, “An em algorithm for wavelet-based image restoration,” *IEEE Trans. on Image Processing.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[69] I. Daubechies, M. Defriese, and C. DeMol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[70] M. Figueiredo and R. Nowak, “A bound optimization approach to wavelet-based image deconvolution,” in *Proc. IEEE Intern. Conf. on Image Process.*, vol. 2, Genoa, Italy, Sep. 2005, pp. 782–785.

[71] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[72] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted norm minimization algorithm," *IEEE Trans. on Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[73] J. Murray and K. Kreutz-Delgado, "An improved focuss-based learning algorithm for solving sparse linear inverse problems," in *Conf. Record of the Thirty-Fifth Asilomar Conf. on Signals, Systems and Computers*, vol. 1, Nagoya, Japan, Nov. 2001, pp. 347–351.

[74] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[75] R. Gribonval and M. Nielsen, "On the strong uniqueness of highly sparse expansions from redundant dictionaries," in *Proc. Int Conf. Independent Component Analysis*, Granada, Spain, Sep. 2004, pp. 201–208.

[76] ——, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, 2007.

[77] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory.*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[78] A. Juditsky and A. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via l_1 minimization," *Math. Program. A. B.*, vol. 127, no. 1, pp. 57–88, Mar. 2011.

[79] A. Serdaroglu, B. Yazici, and K. Kwon, "Optimum source design for detection of heterogeneities in diffuse optical imaging," in *Proc. of SPIE Photonic West*, vol. 6139, San Jose, CA, Jan. 2006, p. 61391A.

[80] A. Joshi, W. Bangerth, and E. Sevick-Muraca, “Non-contact fluorescence optical tomography with scanning patterned illumination,” *Opt. Express*, vol. 14, no. 14, pp. 6516–6534, 2006.

[81] A. Thompson and E. Sevick-Muraca, “Nir fluorescence contrast enhanced imaging with iccd homodyne detection: Measurement precision and accuracy,” *J. Biomed. Opt.*, vol. 8, no. 1, pp. 111–120, 2003.

[82] A. Joshi, W. Bangerth, and E. Sevick-Muraca, “Adaptive finite element modeling of optical fluoresc enhanced tomography,” *Opt. Express*, vol. 12, pp. 5402–5417, 2004.

[83] A. Joshi, W. Bangerth, K. Hwang, J. Rasmussen, and E. Sevick-Muraca, “Plane wave fluorescence tomography with adaptive finite elements,” *Opt. Lett.*, vol. 31, no. 2, pp. 193–195, 2006.

[84] A. Godavarty, A. Thompson, R. Roy, M. Gurfinkel, M. Eppstein, C. Zhang, and E. Sevick-Muraca, “Diagnostic imaging of breast cancer using fluorescence-enhanced optical tomography: phantom studies,” *J. Biomed. Opt.*, vol. 9, no. 3, pp. 488–496, 2004.

[85] V. Lukic, V. Markel, and J. Schotland, “Optical tomography with structured illumination,” *Opt. lett.*, vol. 34, no. 7, pp. 983–985, 2009.

[86] J. Chen, V. Venugopal, F. Lesage, and X. Intes, “Time-resolved diffuse optical tomography with patternedlight illumination and detection,” *Opt. lett.*, vol. 35, no. 13, pp. 188–191, 2010.

[87] A. Olivo and R. Speller, “A coded-aperture technique allowing x-ray phase contrast imaging with conventional sources,” *Appl. Phys. Lett.*, vol. 91, p. 074106, 2007.

[88] H. Weissa, E. Klotza, R. Lindea, G. Rabea, and U. Tiemensa, “Coded aperture imaging with x-rays (flashing tomosynthesis),” *Optica Acta: Int. J. Opt.*, vol. 24, no. 4, pp. 305–325, 1977.

[89] S. Meikle, “An investigation of coded aperture imaging for small animal spect,” *IEEE Trans. Nucl. Sci.*, vol. 48, no. 3, pp. 816–821, Jun. 2001.

[90] E. Caroli, J. Stephen, G. Cocco, L. Natalucci, and A. Spizzichino, “Coded aperture imaging in x- and gamma-ray astronomy,” *Space Sci. Rev.*, vol. 45, no. 3-4, pp. 349–403, 1987.

[91] A. Wagadarikar, R. John, R. Willett, and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging,” *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.

[92] H. Arguello and G. Arce, “Code aperture optimization for spectrally agile compressive imaging,” *JOSA A*, vol. 28, no. 11, pp. 2400–2413, 2011.

[93] R. Marcia and R. Willett, “Compressive coded aperture superresolution image reconstruction,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Las Vegas, NV, May 2008, pp. 833–836.

[94] R. Marcia, Z. Harmany, and R. Willett, “Compressive coded apertures for high-resolution imaging,” in *Proc. 2010 SPIE Photonics Europe*, vol. 7723, Brussels, Belgium, Apr. 2010, pp. 772 304–772 311.

[95] J. Ye, S. Lee, and Y. Bresler, “Exact reconstruction formula for diffuse optical tomography using simultaneous sparse representation,” in *Proc. 2008 IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, Paris, France, May 2008, pp. 1621–1624.

[96] M. Suzen, A. Giannoula, and T. Durduran, “Compressed sensing in diffuse optical tomography,” *Opt. Express.*, vol. 18, no. 23, pp. 23 676–23 690, 2010.

[97] M. Suzen, A. Giannoula, P. Zirak, N. Oliverio, U. Weigel, P. Farzam, and T. Durduran, “Sparse image reconstruction in diffuse optical tomography: an application of compressed sensing,” in *Proc. Biomed. Opt. (BIOMED)*, Miami, FL, Apr. 2010, p. BSuE6.

[98] O. Lee, J. Kim, Y. Bresler, and J. Ye, “Compressive diffuse optical tomography: Non-iterative exact reconstruction using joint sparsity,” *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1129–1142, May 2011.

[99] A. Jin and B. Yazici, “Combined l_1-l_2-norm regularization for fluorescence diffuse optical tomography,” in *IEEE Northeast Bioengineering Conf. (NEBEC)*, Troy, NY, Apr. 2011, pp. 1–2.

[100] ——, “Sparse reconstruction methods for fluorescence diffuse optical tomography,” unpublished.

[101] A. Jin, B. Yazici, A. Ale, and V. Ntziachristos, “Performance evaluation of fluorescence diffuse optical tomography sensing matrix preconditioning using real data,” *Opt. lett.*, vol. 30, no. 20, pp. 4326–4328, 2012.

[102] A. Jin, B. Yazici, and V. Ntziachristos, “Light illumination and detection patterns for fluorescence diffuse optical tomography based on compressive sensing,” *IEEE Trans. Image Process.*, submitted for publication.

[103] B. Alacam and B. Yazici, “Direct reconstruction of pharmacokinetic-rate images of optical fluorophores from nir measurements,” *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1337–1353, Sep. 2009.

[104] A. Tikhonov and V. Arsenin, *Solutions of Ill-posed Problems.* Washington, DC: Winston, 1977.

[105] J. Shewchuk, “An introduction to the conjugate gradient method without the agonizing pain,” Carnegie Mellon University, Pittsburgh, PA, Tech. Rep., 1994.

[106] R. Gribonval, R. Ventura, and P. Vandergheynst, “A simple test to check the optimality of a sparse signal approximation,” *Signal Process.*, vol. 86, no. 3, pp. 496–510, 2006.

[107] H. Vinod, “A survey of ridge regression and related techniques for improvements over ordinary least squares,” *Rev. Economics Statistics*, vol. 60, no. 1, pp. 121–131, 1978.

[108] Q. Zhang, T. Brukilacchio, A. Li, J. Stott, T. Chaves, E. Hillman, T. Wu, M. Chorlton, E. Rafferty, R. Moore, D. Kopans, and D. Boas, “Coregistered tomographic x-ray and optical breast imaging: initial results,” *J Biomed. Opt.*, vol. 10, no. 2, p. 024033, 2005.

[109] A. Li, E. Miller, M. Kilmer, T. Brukilacchio, T. Chaves, J. Stott, Q. Zhang, T. Wu, M. Chorlton, R. Moore, D. Kopans, and D. Boas, “Tomographic optical breast imaging guided by three-dimensional mammography,” *Appl. Opt.*, vol. 42, no. 25, pp. 5181–5190, 2003.

[110] D. Cuccia, F. Bevilacqua, A. Durkin, S. Merritt, B. Tromberg, G. Gulsen, H. Yu, J. Wang, and O. Nalcioglu, “In vivo quantification of optical contrast agent dynamics in rat tumors by use of diffuse optical spectroscopy with magnetic resonance imaging coregistration,” *Appl. Opt.*, vol. 42, no. 16, pp. 2940–50, 2003.

[111] B. Pogue and K. Paulsen, “High-resolution near-infrared tomographic imaging simulations of the rat cranium by use of a priori magnetic resonance imaging structural information,” *Opt. Lett.*, vol. 23, no. 21, pp. 1716–1720, 1998.

[112] Q. Zhu, N. Chen, and S. Kurtzman, “Imaging tumor angiogenesis by use of combined near-infrared diffusive light and ultrasound,” *Opt. Lett.*, vol. 28, no. 5, pp. 337–339, 2003.

[113] Q. Zhu, M. Huang, N. Chen, K. Zarfos, B. Jagjivan, M. Kane, P. Hedge, and S. Kurtzman, “Ultrasound-guided optical tomographic imaging of malignant and benign breast lesions: Initial clinical results of 19 cases,” *Neoplasia*, vol. 5, no. 5, pp. 379–388, 2003.

[114] H. Dehghani, B. Pogue, S. Jiang, B. Brooksby, and K. Paulsen, “Three-dimensional optical tomography: Resolution in small-object imaging,” *Appl. Opt.*, vol. 42, no. 16, pp. 3117–3128, 2003.

[115] D. Boas and A. Dale, “Simulation study of magnetic resonance imagingcguided cortically constrained diffuse optical tomography of human brain function,” *Appl. Opt.*, vol. 44, no. 10, pp. 1957–1968, 2005.

[116] M. Schweiger and S. Arridge, “Optical tomographic reconstruction in a complex head model using a priori region boundary information,” *Appl. Opt.*, vol. 44, no. 11, pp. 2703–2721, 1999.

[117] P. Yalavarthy, B. Pogue, H. Dehghani, and K. Paulsen, “Weight-matrix structured regularization provides optimal generalized least-squares estimate in diffuse optical tomography,” *Med. Phys.*, vol. 34, no. 6, pp. 2085–2098, 2007.

[118] B. Brooksby, S. Jiang, H. Dehghani, B. Pogue, K. Paulsen, J. Weaver, C. Kogel, and S. Poplack, “Combining near-infrared tomography and magnetic resonance imaging to study in vivo breast tissue: implementation of a laplacian-type regularization to incorporate magnetic resonance structure,” *J. Biomed. Opt.*, vol. 10, no. 5, p. 051504, 2005.

[119] S. Davis, H. Dehghani, J. Wang, S. Jiang, B. Pogue, and K. Paulsen, “Image-guided diffuse optical fluorescence tomography implemented with laplacian-type regularization,” *Opt. Express*, vol. 15, no. 7, pp. 4066–4082, 2007.

[120] G. Golub and F. Van-Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.

[121] D. Omidiran and J. Wainwright, “High-dimensional subset recovery in noise: Sparsified measurements without loss of statistical efficiency,” University of California, Berkeley, Berkeley, CA, Tech. Rep., May 2008.

[122] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia, “Compressed sensing performance bounds under poisson noise,” *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3990–4002, Aug. 2010.

[123] Z. Ben-Haim, Y. Eldar, and M. Elad, “Coherence-based performance guarantees for estimating a sparse vector under random noise,” *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[124] K.Schnass and P. Vandergheynst;, “Dictionary preconditioning for greedy algorithms,” *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1994–2002, May 2008.

[125] J. Xu, Y. Pi, and Z. Cao, “Optimized projection matrix for compressive sensing,” *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 43, 2010.

[126] G. Tang and A. Nehorai, “Performance analysis of sparse recovery based on constrained minimal singular values,” *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5734–5745, Dec. 2011.

[127] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[128] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[129] K. Seung-Jean, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, “An interior-point method for large-scale l_1-regularized least squares,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[130] E. Candes and J. Romberg, “l_1-magic: Recovery of sparse signals via convex programming,” Standard University, Stanford, CA, Tech. Rep., Oct. 2005.

[131] J. Barzilai and J. Borwein, “Two point step size gradient methods,” *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.

[132] N. Deliolanis, T. Lasser, D. Hyde, A. Soubret, J. Ripoll, and V. Ntziachristos, “Free-space fluorescence molecular tomography utilizing 360 degrees geometry projections,” *Opt. Lett.*, vol. 32, no. 4, pp. 382–384, 2007.

[133] L.Zelnik-Manor, K.Rosenblum, and Y. Eldar, “Sensing matrix optimization for block-sparse decoding,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.

APPENDIX A

Calculation of the Combined l_2-l_p norm Regularization Coefficients

The penalty function $r_{com}(x)$ is given as

$$r_{com}(x) = \begin{cases} c_1x^2, & |x| \leq \kappa_1; \\ c_2|x|^p + c_3, & \kappa_1 \leq |x| \leq \kappa_2; \\ c_4x^2 + c_5|x|^p + c_6, & |x| > \kappa_2. \end{cases} \tag{A.1}$$

$r_{com}(x)$ is continuous and differentiable $\forall x \in \mathbb{R}$, which consists of three parts. Let $r^{(1)}_{com}(x)$, $r^{(2)}_{com}(x)$ and $r^{(3)}_{com}(x)$ denote the three parts respectively:

$$r^{(1)}_{com}(x) = c_1x^2, \qquad |x| \leq \kappa_1; \tag{A.2}$$

$$r^{(2)}_{com}(x) = c_2|x| + c_3, \quad \kappa_1 \leq |x| \leq \kappa_2; \tag{A.3}$$

$$r^{(3)}_{com}(x)c_4x^2 + c_5|x| + c_6, \qquad |x| > \kappa_2. \tag{A.4}$$

$r_{com}(x)$ is continuous and differentiable at the joint points $x = \kappa_1$, κ_2:

$$r^{(1)}_{com}(\kappa_1) = r^{(2)}_{com}(\kappa_2) \tag{A.5}$$

$$\frac{dr^{(1)}_{com}(x)}{dx}|_{x=\kappa_1} = \frac{dr^{(2)}_{com}(x)}{dx}|_{x=\kappa_1} \tag{A.6}$$

$$r^{(2)}_{com}(\kappa_2) = r^{(3)}_{com}(\kappa_2), \tag{A.7}$$

$$\frac{r^{(2)}_{com}(x)}{dx}|_{x=\kappa_2} = \frac{dr^{(3)}_{com}(x)}{dx}|_{x=\kappa_2}. \tag{A.8}$$

The penalty function $r_{com}(x)$ is impose smaller penalty than the quadratic penalty when $x \in [\alpha, \beta]$, and larger penalty when $x \notin [\alpha, \beta]$. Since both $r_{com}(x)$ and $r_2(x)$ are continuous functions, thus

$$r_{com}(\alpha) = r_2(\alpha) = \alpha^2. \tag{A.9}$$

Similarly, we have

$$r_{com}(\beta) = r_2(\beta) = \beta^2 \tag{A.10}$$

From (A.5)-(A.10), we have six equations to solve for six unknown coefficients c_1, c_2, c_3, c_4, c_5 and c_6, and the solution is given as:

$$c_1 = \frac{p\alpha^2\kappa_1^{p-2}}{2\alpha^p + (p-2)\kappa_1^p}, \tag{A.11}$$

$$c_2 = \frac{2\alpha^2}{2\alpha^p + (p-2)\kappa_1^p}, \tag{A.12}$$

$$c_3 = \frac{(p-2)\alpha^2\kappa_1^p}{2\alpha^p + (p-2)\kappa_1^p}, \tag{A.13}$$

$$c_4 = \frac{p\kappa_2^{p-1}(\beta^2 - c_2\beta^p - c_3)}{p\kappa_2^{p-1}\beta^2 + (2-p)\kappa_2^{p+1} - 2\kappa_2\beta^p}, \tag{A.14}$$

$$c_5 = c_2 - \frac{2\kappa_2^{2-p}c_4}{p}, \tag{A.15}$$

$$c_6 = \beta^2 - \beta^2 c_4 - \beta^p c_5. \tag{A.16}$$

APPENDIX B

Upper Bounds for Cumulative Coherence

We recall that the definition of the cumulative coherence,

$$\mathcal{M}_1(k,\boldsymbol{A}) = \max_p \max_{|Q|=k, p\notin Q} \sum_{q\in Q} \frac{|<\boldsymbol{a}_p,\boldsymbol{a}_q>|}{\|\boldsymbol{a}_p\|_2\|\boldsymbol{a}_q\|_2} = \max_p \max_{|Q|=k, p\notin Q} \sum_{q\in Q} |r_{A_{p,q}}|. \tag{B.1}$$

Let

$$R(p,Q_k) = \sum_{|Q_k|=k, p\notin Q_k, q\in Q_k} |r_{A_{p,q}}|, \tag{B.2}$$

and we assume that

$$R(p_0,Q_{k_0}) = \max_{p,Q_k} R(p,Q_k) = \mathcal{M}_1(k,\boldsymbol{A}), \quad \text{such that} \quad |Q_k| = k \text{ and } p\notin Q_k. \tag{B.3}$$

We note that

$$R(p_0,Q_{k_0}) = \sum_{q\in Q_{k_0}} r_{A_{p_0,q}} = \sum_{q\in Q_{k_0}} r_{\phi_{p_0,q}} r_{g_{p_0,q}} \le \frac{1}{2}\sum_{q\in Q_{k_0}} (r^2_{\phi_{p_0,q}} + r^2_{g_{p_0,q}}). \tag{B.4}$$

Let $\tilde{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{G}}$ denote the $\boldsymbol{\Phi}$ and $\boldsymbol{G}$ matrices with each column normalized to unit respectively. $\tilde{\boldsymbol{\Phi}}_{Q_{k_0}}$ is the $N_s \times k$ submatrix of $\tilde{\boldsymbol{\Phi}}$ composed the columns in Q_{k_0}, and $\tilde{\boldsymbol{G}}_{Q_{k_0}}$ is the $N_d \times k$ submatrix of $\tilde{\boldsymbol{G}}$ composed the columns in Q_{k_0}. Thus, we have

$$\begin{aligned} \sum_{q\in Q_{k_0}} (r^2_{\phi_{p_0,q}} + r^2_{g_{p_0,q}}) &= \|\tilde{\boldsymbol{\Phi}}^T_{Q_{k_0}} \tilde{\boldsymbol{\phi}}_{p_0}\|_2^2 + \|\tilde{\boldsymbol{G}}^T_{Q_{k_0}} \tilde{\boldsymbol{\phi}}_{p_0}\|_2^2 \\ &\le \lambda_{\max_{\phi_k}} \|\tilde{\boldsymbol{\phi}}_{p_0}\|_2^2 + \lambda_{\max_{g_k}} \|\tilde{\boldsymbol{g}}_{p_0}\|_2^2 = \lambda_{\max_{\phi_k}} + \lambda_{\max_{g_k}}, \end{aligned} \tag{B.5}$$

where $\lambda_{\max_{\phi_k}}$ is the maximum eigenvalue of the $k \times k$ matrix $\boldsymbol{H}_\phi = \tilde{\boldsymbol{\Phi}}^T_{Q_{k_0}} \tilde{\boldsymbol{\Phi}}_{Q_{k_0}}$, and $\lambda_{\max_{g_k}}$ is the maximum eigenvalue of the $k \times k$ matrix $\boldsymbol{H}_g = \tilde{\boldsymbol{G}}^T_{Q_{k_0}} \tilde{\boldsymbol{G}}_{Q_{k_0}}$. The maximum value of $\lambda_{\max_{\phi_k}}$ and $\lambda_{\max_{g_k}}$ is k. From (B.3)-(B.5),

$$\mathcal{M}_1(k,\boldsymbol{A}) \le \frac{1}{2}(\lambda_{\max_{\phi_k}} + \lambda_{\max_{g_k}}) \tag{B.6}$$

Let $\boldsymbol{H}_\phi = \tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}}$ and $\boldsymbol{H}_g = \tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}}$ denote the Gram type matrix, and let $\lambda_{\phi_1} \geq \lambda_{\phi_2} \geq \ ... \lambda_{\phi_{N_s}} \geq 0$ denote the nonzero eigenvalues of $\boldsymbol{H}_\phi$, $\lambda_{g_1} \geq \lambda_{g_2} \geq \ ... \lambda_{g_{N_d}} \geq 0$ denote the nonzero eigenvalues of $\boldsymbol{H}_g$, we have

$$\|\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2 = \sum_{i=1}^{N_s} \lambda_{\phi_i}^2 - N \quad \text{and} \quad \|\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}} - \boldsymbol{I}_N\|_F^2 = \sum_{i=1}^{N_d} \lambda_{g_i}^2 - N. \tag{B.7}$$

We derive the ratio r_1 between $\lambda_{\max_{\phi_k}}$ and $\|\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2$ as follows,

$$\begin{aligned} r_1 \ &= \ \frac{\lambda_{\max_{\phi_k}}}{\|\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2} = \frac{\lambda_{\max_{\phi_k}}}{\sum_{i=1}^{N_s} \lambda_{\phi_i}^2 - N} \\ &\leq \ \frac{\lambda_{\max_{\phi_k}}}{\frac{N^2}{N_s} - N} \leq \frac{k}{\frac{N^2}{N_s} - N}. \end{aligned} \tag{B.8}$$

The first inequality is because if all the eigenvalues λ_{ϕ_i} $(i = 1, ..., N_s)$ are equal to each other $(\lambda_{\phi_i} = \frac{N}{N_s})$, their square summation is minimum. The second inequality is because that $\lambda_{\max_{\phi_k}}$ has maximum value k.

However, for large k values, the upper bound derived in (B.8) seem to be too large for the ratio r_1. In this case, we notice that $\tilde{\boldsymbol{\Phi}}_{Q_{k_0}}$ is a submatrix of $\tilde{\boldsymbol{\Phi}}$, thus $\lambda_{\phi_1} \geq \lambda_{\max_{\phi_k}}$, we have

$$\begin{aligned} r_1 \ &= \ \frac{\lambda_{\max_{\phi_k}}}{\|\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2} = \frac{\lambda_{\max_{\phi_k}}}{\sum_{i=1}^{N_s} \lambda_{\phi_i}^2 - N} \\ &\leq \ \frac{\lambda_{\phi_1}}{\sum_{i}^{N_s} \lambda_{\phi_i}^2 - N} \leq \frac{\lambda_{\phi_1}}{\lambda_{\phi_1}^2 + \frac{(N - \lambda_{\phi_1})^2}{N_s - 1} - N} \leq \frac{N_s - 1}{2(\sqrt{N_s N^2 - N_s^2 N + N_s N} - N)}. \end{aligned} \tag{B.9}$$

From both (B.8) and (B.9),

$$r_1 = \frac{\lambda_{\max_{\phi_k}}}{\|\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} - \boldsymbol{I}_N\|_F^2} \leq \min\{\frac{k N_s}{N^2 - N N_s}, \frac{N_s - 1}{2(\sqrt{N_s N^2 - N_s^2 N + N_s N} - N)}\}. \tag{B.10}$$

Similarly, we derive the ratio between $\lambda_{\max_{g_k}}$ and $\|\tilde{\boldsymbol{G}}^T\tilde{\boldsymbol{G}}-\boldsymbol{I}_N\|_F^2$ is,

$$r_2=\frac{\lambda_{\max_{g_k}}}{\|\tilde{\boldsymbol{G}}^T\tilde{\boldsymbol{G}}-\boldsymbol{I}_N\|_F^2}\leq\min\{\frac{kN_d}{N^2-NN_d},\frac{N_d-1}{2(\sqrt{N_dN^2-N_d^2N+N_dN}-N)}\}. \tag{B.11}$$

Since $\mathcal{M}_1(k,\boldsymbol{A})\leq\frac{1}{2}(\lambda_{\max_{\phi_k}}+\lambda_{\max_{g_k}})$,

$$\begin{aligned}\mathcal{M}_1(k,\boldsymbol{A}) &\leq \frac{r_1}{2}\|\tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}}-\boldsymbol{I}_N\|_F^2+\frac{r_2}{2}\|\tilde{\boldsymbol{G}}^T\tilde{\boldsymbol{G}}-\boldsymbol{I}_N\|_F^2\\ &\leq \frac{\max\{\bar{r}_1,\bar{r}_2\}}{2}(\|\tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}}-\boldsymbol{I}_N\|_F^2+\|\tilde{\boldsymbol{G}}^T\tilde{\boldsymbol{G}}-\boldsymbol{I}_N\|_F^2), \qquad \text{(B.12)}\end{aligned}$$

where

$$\bar{r}_1=\min\{\frac{kN_s}{N^2-NN_s},\frac{N_s-1}{2(\sqrt{N_sN^2-N_s^2N+N_sN}-N)}\}, \tag{B.13}$$

and

$$\bar{r}_2=\min\{\frac{kN_d}{N^2-NN_d},\frac{N_d-1}{2(\sqrt{N_dN^2-N_d^2N+N_dN}-N)}\}. \tag{B.14}$$

APPENDIX C

Equivalence of the Normalized Forward Sensing Matrix

Consider normalizing every column of the matrices $\boldsymbol{\Phi}$ and $\boldsymbol{G}$ to unit,

$$\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{D}_\phi^{-1} \quad \text{and} \quad \tilde{\boldsymbol{G}} = \boldsymbol{G}\boldsymbol{D}_g^{-1} \tag{C.1}$$

where $\boldsymbol{D}_\phi$ and $\boldsymbol{D}_g$ are two diagonal matrices, with diagonal entries equal to the l_2-norms of the columns of $\boldsymbol{\Phi}$ and $\boldsymbol{G}$, i.e.,

$$d_{\phi_{k,k}} = \|\boldsymbol{\phi}_k\|_2 \quad \text{and} \quad d_{g_{k,k}} = \|\boldsymbol{g}_k\|_2, \quad \text{for} \quad k = 1, 2, ..., N.$$

Let $\tilde{\boldsymbol{\phi}}_k = \frac{1}{d_{\phi_{k,k}}}\boldsymbol{\phi}_k$ denote the kth column of $\tilde{\boldsymbol{\Phi}}$, and $\tilde{\boldsymbol{g}}_k = \frac{1}{d_{g_{k,k}}}\boldsymbol{g}_k$ denote the k^{th} column of the matrix $\tilde{\boldsymbol{G}}$, the forward sensing matrix $\boldsymbol{A}$ can be reformulated as

$$\begin{aligned} \boldsymbol{A} &= [(d_{\phi_{1,1}} d_{g_{1,1}})\tilde{\boldsymbol{\phi}}_1 \otimes \tilde{\boldsymbol{g}}_1, ..., (d_{\phi_{N,N}} d_{g_{N,N}})\tilde{\boldsymbol{\phi}}_N \otimes \tilde{\boldsymbol{g}}_N] \\ &= \tilde{\boldsymbol{A}}\boldsymbol{D}_\phi\boldsymbol{D}_g \end{aligned} \tag{C.2}$$

where $\tilde{\boldsymbol{A}} = [\tilde{\boldsymbol{\phi}}_1 \otimes \tilde{\boldsymbol{g}}_1, ..., \tilde{\boldsymbol{\phi}}_N \otimes \tilde{\boldsymbol{g}}_N]$ is the forward sensing matrix with all the columns normalized to unit.

The objective is to recovery the fluorophore yield $\boldsymbol{x}$ from the linear measurement system,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{C.3}$$

which can be reformulated as

$$\boldsymbol{y} = \tilde{\boldsymbol{A}}\boldsymbol{D}_\phi\boldsymbol{D}_g\boldsymbol{x} + \boldsymbol{\epsilon} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} + \boldsymbol{\epsilon}, \tag{C.4}$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{D}_\phi\boldsymbol{D}_g\boldsymbol{x}$. We recover $\tilde{\boldsymbol{x}}$ instead of $\boldsymbol{x}$ from the equivalent normalized forward sensing matrix $\tilde{\boldsymbol{A}}$. Note that $\tilde{\boldsymbol{x}}$ has the same sparsity level as $\boldsymbol{x}$. The original fluorophore yield $\boldsymbol{x}$ can be easily derived by

$$\boldsymbol{x} = \boldsymbol{D}_{\phi}^{-1} \boldsymbol{D}_{g}^{-1} \tilde{\boldsymbol{x}}. \tag{C.5}$$