

**Statistical Modeling and Structured Regularization for
Fluorescence Molecular Tomography**

A Dissertation Presented

by

Damon Eliot Hyde

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Electrical Engineering

Northeastern University

Boston, Massachusetts

December 2008

NORTHEASTERN UNIVERSITY
Graduate School of Engineering

Dissertation Title: Statistical Modeling and Structured Regularization for Fluorescence Molecular Tomography

Author: Damon Eliot Hyde

Department: Electrical and Computer Engineering

Approved for Dissertation Requirement for the Doctor of Philosophy Degree

Prof. Dana Brooks

Date

Prof. Eric Miller

Date

Prof. Vasilis Ntziachristos

Date

Prof. Mark Niedre

Date

Dr. Ali Abur, Department Chair

Date

Graduate School Notified of Acceptance:

Dr. Yaman Yener, Graduate School Director

Date

Hyde, Damon Eliot (PhD E.E.)

Statistical Modeling and Structured Regularization for Fluorescence Molecular Tomography

Thesis directed by Prof. Dana Brooks, Prof. Eric Miller, and Prof. Vasilis Ntziachristos

Fluorescence molecular tomography (FMT) is an optical imaging technique that uses near infrared light to localize and quantify *in vivo* distributions of fluorescent probes targeting biochemical markers such as genes, proteins, and enzymes. In this thesis, we examine three aspects of the FMT reconstruction problem: statistical data modeling in the context of normalized fluorescence imaging, methods for the use of prior structural information arising from multi-modal FMT-CT imaging, and techniques to compensate for errors in that prior information. We derive a probabilistic model for normalized fluorescence data and use this model as the basis for reconstruction. This eliminates errors and human biases introduced by manual data thresholding and is shown to yield improved reconstructions with greater consistency. To improve upon the resolution limits of stand-alone FMT, we examine modeling and regularization that incorporates structural prior information available from data acquired by a complementary imaging modality such as CT or MRI. We show that improved diffusion forward models using average tissue optical properties can subsequently result in improved reconstructions. A two step inversion approach is then presented, using the solution to an anatomically defined low dimensional problem as the basis of a spatially varying regularization term for the full resolution problem. Results are presented for both simulated and *in vivo* data, in the context of imaging a mouse model of Alzheimer’s disease. Such diffuse targets are difficult to reconstruct with stand alone approaches, thus highlighting the utility of the multimodal approach. Results are correlated with post mortem fluorescence measurements, and show a high degree of correlation between reconstruction intensity and observed fluorescence. Finally, two methods are presented to address situations where the prior information and underlying fluorescence share similar, but not identical, structure. The first uses differential equations to derive a Gaussian prior model for the fluorescence image. The incorporation of boundary conditions between anatomical regions allows information to cross their boundaries, and can help to compensate for boundary misplacement in the prior. The second approach uses the sparsity inducing properties of 1-norm minimization to localize the boundary within an uncertainty region around its initial position. Both approaches are tested using a range of 2-D simulated experiments.

Dedication

To my parents.

Acknowledgements

Prof Brooks

Prof Miller

Vasilis

Josh Dunham

Mark Niedre

Nikos Deliolanis

Ruben DeKleine

Claudio Vinegoni

Antoine Soubret

Gordon Turner

Thomas Krucker

Ralf Schulz

Contents

Chapter

1	Introduction	1
1.1	Statement of Problem	1
1.2	Thesis Contributions	2
1.3	Optical Medical Imaging	6
1.3.1	Diffuse Optical Tomography	6
1.3.2	Radiative Transport	7
1.3.3	Diffusion Approximation	8
1.3.4	Born Approximation	9
1.3.5	Linearization and Formation of Inverse Problem	10
1.4	Fluorescence Diffusion Imaging	12
1.4.1	Normalized Born Ratio	13
1.5	Non-contact Imaging	14
1.5.1	Surface Reconstruction	16
1.6	Fluorescence Molecular Imaging Systems	16
1.6.1	Planar FMT System	17
1.6.2	Rotational Imaging System	18
2	Inverse Problems and Solution Methods	22
2.1	Introduction to Inverse Problems	22
2.2	Solution Methods	26

2.2.1	Algebraic Reconstruction Technique	26
2.2.2	Conjugate Gradient	28
2.2.3	LSQR	30
2.2.4	Modified Residual Norm Steepest Descent	31
2.3	Regularization	32
2.3.1	Ill-Posedness	32
2.3.2	L-Curve Analysis	35
3	Statistical Modeling of the Born Ratio	40
3.1	Introduction	40
3.1.1	Signal Modeling	41
3.1.2	Density Derivation	45
3.1.3	Density Conditioning	47
3.1.4	Maximum Likelihood Solution	48
3.1.5	Results and Discussion	50
3.1.6	Phantoms in Intralipid	53
3.1.7	Euthanized Mouse	55
3.1.8	In-Vivo Mouse	58
3.2	Conclusions	58
4	Effects of Parameter Selection on Finite Element Modeling for Normalized Fluorescence Imaging	63
4.1	Introduction	63
4.2	Methods	65
4.3	Results	68
4.4	Acknowledgments	71
5	MultiModality Imaging	72
5.1	Introduction	72

5.2	Structural Priors	77
5.2.1	Low-Dimensional Parameterized Inverse	77
5.2.2	Regularization Construction	79
5.2.3	Summary of Approach	81
5.3	Experimental Methods	81
5.3.1	Anatomic Subsegmentation	82
5.3.2	Mouse Brain Simulation Studies	85
5.3.3	In Vivo Imaging Study	86
5.3.4	Acknowledgements	90
5.4	Results	90
5.4.1	Partial Volume Labeling	90
5.4.2	Simulations	91
5.4.3	In-Vivo FMT-CT Imaging of Amyloid- β Plaques	93
5.5	Optimal Parameter Estimation	95
5.6	Discussion and Conclusions	98
6	Robust use of Structural Information	109
6.1	Introduction	109
6.2	Differential Equation Based Covariance Models	111
6.2.1	From Differential Equation to Covariance Matrix	112
6.2.2	An Example Implementation	113
6.2.3	Effects of Parameter Selection on Covariance	116
6.2.4	Results	121
6.3	1-Norm Based Floating Boundaries	122
6.3.1	Limited Angle Radon	127
6.3.2	Results	128
6.4	Conclusions	129

7	Conclusions	133
----------	--------------------	------------

Appendix

A	Detailed Derivations of Density Approximation	157
A.1	Approximation of terms involving Φ	157
A.2	Elimination of Low Magnitude Term	159
A.3	Deriving the Gradient of the Cost Function	162
A.3.1	Primary Derivations	162
A.3.2	Derivation of one of the gradient components	164

Tables

Table

3.1	Locations and Sizes of Reconstructed Objects Centroid and Maximum Dimension are given in (x,y,z) coordinates except for the subcutaneous tube in the euthanized mouse, where width, length, depth of the tube are given, as it is not parallel with any of the three primary axes. Measurements on the reconstructions are computed as described in the text. All size measurements are given in cm.	54
4.1	Modeled and Published values for μ_a	66
4.2	Modeled and Published values for μ'_s	67

Figures

Figure

- 1.1 **Diagram of Planar FMT Imaging System** Laser light from the source laser is coupled into the imaging volume through an optical switch and 46 optical fibers arranged in a grid pattern at the back of the imaging chamber. Data collected is done in parallel using a CCD camera and virtual detectors defined as small integrated regions of the CCD. 19
- 1.2 **Rotational FMT System and Localization of Animal Surface** a) Schematic of data collection system. Source fiber driven by a 650nm diode laser was mounted on translation stages for focusing and localization on animal surface. Collection of data was done in a transmission geometry using an electrically cooled CCD camera and appropriate filters for wavelength selection. A representative excitation channel image is shown. b) White light image of mouse head. c) Corresponding silhouette image used for surface localization. 360 such images were used for the surface generation algorithm. d) Summation of normalized (fluorescence/excitation) data across all source locations for this projection angle. e) Resolved surface location, showing source (red) and detector (blue) patterns mathematically projected onto surface for a single projection angle. 21

2.1	Example L-Curve Example of an L-Curve, computed from a non-contact FMT problem. This curve displays the characteristic L shape, although it is not as clearly defined as in some inverse problems. Each point plots the residual norm (X-axis) and solution norm (Y-axis) for a particular regularization parameter λ . As λ increases, so does the norm of the residual, while the norm of the solution decreases. Thus the plot is traversed from lower right to upper left as λ increases. The optimal value of lambda is chosen to be at the corner of the curve, as indicated by the red point in the plot.	36
2.2	L-Curve for a non-negatively constrained problem: Computing an L-Curve prior to truncation of solutions results in the plot denoted with blue X's. If the same L-Curve approach is used on the solutions after truncation, the plot denoted with red squares results. Note that the inflection point of the red curve is as the same point as the corner of the standard L-Curve. This allows for automated selection of regularization parameters when using non-constrained methods to solve constrained problems. . .	39
3.1	Fit of Noise Model to Experimental Data Scatter plot of experimentally collected noise data, along with a regression line from the model in (3.3). The R^2 value for the red fit line is 0.86. The green line shows the model as predicted based solely on published camera parameters. In order to better display the results at all intensity levels, the plot shows standard deviation plotted against the square root of the mean.	43

3.2	Dual Tube Phantom: (a) Reflectance Fluorescence image (b) Manual Threshold: Coronal slice through tube @ $z=0.42\text{cm}$ (c) Manual Threshold: Coronal slice through tube @ $z = 1.25\text{cm}$ (d) Flat (Photographic) Image (e) Fixed Point: Coronal Slice through $z=0.42\text{cm}$ (f) Fixed Point: Coronal Slice through $z=1.25\text{cm}$ In all images, $z = 0\text{cm}$ corresponds to the source side of the slab, with a total slab thickness of 1.3cm . The red box in the image denotes the boundary of the solution space, and the orientation of the coordinate system specified in 3.2(b) is shared by all images.	56
3.3	Euthanized Mouse: (a) Fluorescence Reflectance (b) Manual Threshold: Coronal slice through esophageal tube (c) Manual Threshold: Coronal slice through subcutaneous tube. (d) Flat (Photographic) Image (e) Fixed Point: Coronal slice through esophageal tube (f) Fixed Point: Coronal slice through subcutaneous tube. In all images, $z = 0\text{cm}$ corresponds to the source side of the slab, with a total slab thickness of 1.3cm . The red box denotes the boundary of the solution space.	59
3.4	In-Vivo experiment. (a) Flat (Photographic) Image. Arrow indicates location of implanted tumor. (b) Manually Selected Threshold (c) Fixed Point Iteration. In both reconstructions, the red box denotes the boundary of the solution space.	60
4.1	a) Rendering of FEM geometry b) Rendering showing exterior surface, lung surface, and location of all four fluorescing inclusions used in combination to generate simulated data measurements.	66
4.2	Absolute 2-norm Error using a) standard and b) <i>a priori</i> inversion. Results show consistent improvement when using mid-range and matched parameter models. Error levels for the full lung inclusion incorporating <i>a priori</i> structure in the inverse problem are not shown to better display other results.	69

4.3	Relative 2-norm Error using a) standard and b) apriori inversion. Note that relative penalty for using the incorrect model is significantly higher when using <i>apriori</i> inversion techniques.	70
4.4	a) True full-lung image. b-e) Reconstructions of inclusion using: b) Matched model c) Mismatched Model d) Mid-range model e) Homogeneous Model	71
5.1	Partial Volume Labeling Construction of partial volume labeling matrix. Here, voxel I lies 50% within region 1 and 50% within region 2, while voxel II lies 25% within region 1, 25% within region 2, and 50% within region 3. This is reflected in the rows of the matrix C , each of which is of unit 1-norm.	79
5.2	CT-FMT Registration and Modeling Geometry a) A 2D slice of the CT data, showing labeled brain (red) and cortical (blue) regions. Dotted red lines indicate where regions were subdivided prior to the inversion process. b) An isosurface from the CT image (gray) superimposed with surface computed by FMT volume carving algorithm (yellow). This alignment allows for use of CT information in the FMT problem. c) Alignment between modalities allows the brain (red) and cortical (blue) regions to be appropriately located within the surface computed by the FMT system (yellow). d-e) Contour information was extracted from the segmentation information and used to construct surface meshes prior to generation of a tetrahedral mesh for obtaining solutions to the diffusion equation using the finite element method (FEM).	83

5.3	Anatomic segmentation and subsegmentation. a) The original segmentation separated the brain and brain cortex segments (red and blue segments, respectively), from the remaining soft tissue. Subsegmentation of the cortex divided it into three equal angle subsegments, as denoted by the white dashed lines. b) Plot of number of cortical subsegments against resulting average angle between segments, with (triangle) and without (square) subdivision of the central brain segment. c) Average angle divided by total number of subsegments plotted against number of cortical subsegments. Maximum is seen when cortex is divided into three subsegments.	84
5.4	System Diagram and Internal Head Geometry a) Noncontact FMT imaging system diagram. b) Surface reconstructed with FMT system (yellow) overlaid with brain (red) and cortical (blue) segments obtained from CT structural image.	86
5.5	Effects of Partial Volume Labeling: a) Original Image b-c) Reconstructions without partial volume labeling d-e) Reconstructions employing partial volume labeling.	91
5.6	Simulated Imaging Reconstructions Ground truth is shown in the first column, while results using Tikhonov regularization with the identity matrix are in the second column. The simple <i>a priori</i> technique in the third column restricts the solution to either brain segment. The fourth and fifth columns show reconstructions using the techniques developed in Section 3, with and without explicitly imposing a mean value in the image prior model, respectively. Overlaid numerics indicate relative 2-norm error with respect to ground truth.	104

- 5.7 In Vivo Imaging Reconstructions** A control animal is shown in the first row, while APP23 transgenic animals exhibiting amyloid- β plaques in the cortical segment (ages 26 and 28 months) are shown in the second and third rows. The first column shows reconstructions using Tikhonov regularization with the identity matrix, while the second columns using a simple *a priori* technique which favors solutions in the brain region. The third and forth columns show reconstructions using the techniques developed in Section 3, with and without explicitly imposing a mean value in the image prior model, respectively. 105
- 5.8 In-Vivo vs Ex-Vivo Correlation** a) Correlation between *in vivo* reconstructed activity and normalized fluorescence intensity measured *ex vivo* (x-axis) ($R^2=0.9587$) . b) Comparison of mean input FMT data for a single projection (x-axis) with resulting FMT reconstructed intensity ($R^2=0.7422$) . The single projection was chosen with the camera positioned directly above the dorsal side of the head. c) Mean input FMT signal (x-axis) as in (b), correlated with measured ex vivo fluorescence activity ($R^2=0.8712$) 106

5.9	In Vivo vs. Ex Vivo Imaging	Ex vivo versus <i>in vivo</i> FMT imaging comparison for 13 month old C57B/6 control mouse (First row) a 17 month old APP23 tg mouse (Second row) and a 26 month old APP23 tg mouse (Third row). (a) The first column shows full brain images in the excitation channel using a planar reflectance imaging system. The red dotted line denotes the approximate location corresponding to the slice shown in subsequent columns. (b) Planar reflectance images of normalized fluorescence from a single slice are then presented in the second column. (c) The third column presents planar images at the excitation wavelength (top) and FMT reconstructions overlaid on normalized planar fluorescence images (bottom). (d) <i>in vivo</i> multi-modal FMT reconstructions are shown in the final column for a slice corresponding to the same location as the <i>ex vivo</i> images, overlaid on a representative CT slice. All FMT reconstructions are scaled to the same colorbar.	107
5.10	Confocal Imaging	Representative confocal images of thick brain slices from a 13 month old C57B/6 mouse and two APP23 transgenic mice, at ages 23 and 27 months. In each image, amyloid- β plaques appear as bright areas.	108
6.1	Construction of Differential Regularizer	Shown here is a 3×3 voxel region about the voxel x_5 . Here, the voxels $\{x_1, x_4, x_7\}$ lie within region Ω_1 , while the remaining voxels lie within region Ω_2 . The mean values for each region are denoted m_{Ω_1} and m_{Ω_2} , respectively. To construct the row of the differential operator corresponding to x_5 , a 4-neighborhood is employed. The differential operator within Ω_2 results in the first equation written above, while the boundary condition results in the second. The final equation to be satisfied at x_5 is then the sum of these two equations.	115

- 6.2 Change in Covariance Structure as a Function of β :** a) Plot of covariance of voxel 20 with all other voxels, for a range of β values, given $\gamma^2 = 0.5$. b) Plots of the relative change in covariance between voxel 20 (right hand region), and all other voxels, as a function of β . Plots are normalized such that their maximum value is equal to 1. Note that plots are clustered into two groups. The lower, exponentially increasing group, corresponds to the covariance between voxel 20 and voxels within the alternate (left-hand) region. The upper plot shows covariance between voxel 20 and all other voxels within the right-hand region. Note that most voxels have a relative decrease in covariance as β is increased. However several plots, corresponding to those voxels close to the boundary, show a decrease followed by an increase in covariance as β is increased. 118
- 6.3 Change in Covariance Structure as a Function of γ^2 :** Covariance between Voxel 20 and all others for a range of γ^2 values. Note that as γ^2 increases, the covariance plot becomes a double sided exponential, 120
- 6.4 Reconstructions of Simulated Fluorescence within an Artificial 2D Geometry:** a) Fluorescence Distribution used for Data Generation. Red line denotes boundary of lung region b) Reconstruction using Tikhonov Regularization with the Identity Matrix. c) Reconstruction using Laplacian Regularizer applied to each region independently. d) Reconstruction using Differential Equation Based Regularization Technique. 123
- 6.5 Reconstructions using Differential Regularizer on a Complex Target:** a) Fluorescence Distribution used for Data Generation. b) Reconstruction using Tikhonov Regularization with the Identity Matrix. c) Reconstruction using Laplacian Regularizer applied to each region independently. d) Reconstruction using Differential Equation Based Regularization Technique. 124

- 6.6 Diagram of 1-norm Boundary Regularizer** At each of a number of points along the boundary denoted by the heavy black line, the normal vector was computed. Using this direction, and stepping some distance Δ in either direction from the boundary, the integral of the 1-norm of the derivative of the image along that line was computed. The summation of all such 1-norms was then used as a penalty term in the inversion process. 131
- 6.7 The 1-norm and Sparse Optimization** a) By minimizing the 1-norm when solving a system subject to some underdetermined linear constraint $\mathbf{W}\mathbf{x} = \mathbf{u}$, the minimum norm solution which intersects the constraint plane will be a sparse solution, with high probability. b) If the linear constraint is overdetermined, an approach minimizing the norm $\|\mathbf{W}\mathbf{x} - \mathbf{u}\|$ needs to be taken, which likely results in a non-sparse solution. 132
- 6.8 Results Using Boundary Normal 1-Norm Regularization:** a) Fluorescence Distribution used for Data Generation. b) Reconstruction using Tikhonov Regularization with the Identity Matrix. c) Reconstruction using Laplacian Regularization on each region independently. d-f) Reconstructions using 1-norm regularization. d) $\lambda = 0$ e) $\lambda = 0.1$ f) $\lambda = 1$. . . 132
- A.1** Plot of the right hand side of (A.21). Those points with magnitude less than one have been set equal to -1 to give a better idea of the shape of the region in question. 162

Chapter 1

Introduction

1.1 Statement of Problem

Fluorescence based molecular imaging has been an increasingly useful tool in the advancement of biological science, enabling improved resolution of functional activity at levels down to the nanometer scale [1,2]. Fluorescent dyes with high molecular specificity have been designed to target a wide array of biological activity, from protein synthesis to gene and enzyme activity. This capability has enabled researchers to advance biological and medical knowledge, led to a greater understanding of disease pathologies, and assisted in the development of new treatments and cures.

Fluorescence molecular tomography (FMT) takes the advances in probe technology originally pioneered for the field of microscopy and applies them to the problem of *in vivo* imaging in whole animals, thereby allowing studies which would previously have been difficult if not impossible to accomplish [3–17]. While planar epi-illumination systems have been used with fluorescent probes to provide *in vivo* imaging, these methods are limited by the two dimensional nature of the resulting images. The images collected are projections of the three dimensional fluorescence distribution onto the 2D plane of the camera, blurred by the diffusive nature of tissue. Inhomogeneities in the optical properties of the tissue mean that this blurring is variable, and not easily compensated for the way one might compensate for Gaussian blurring in an astronomical image [18]. By using physical models of light diffusion to form and solve an inverse problem, FMT is able

to directly image the three dimensional distributions of fluorescence, resulting in images with greater quantification than those obtained with planar systems. Given its capacity to image whole animals *in vivo*, FMT systems can additionally be used to implement lateral studies, with a single animal imaged at multiple dates, where previously that animal would have been sacrificed to perform a single imaging study. By following animals through the progression of a disease, information can be gathered which is unavailable when only a single imaging session can be performed.

Mathematically, the propagation of light through tissue more than a few millimeters thick is often modeled by the diffusion approximation [19]. This arises as an approximation to the radiative transport equation when assuming isotropic scattering [20]. For the case of fluorescence imaging, a full model needs to use the diffusion approximation twice [21]. The first instance is to describe the diffusion of light at the excitation wavelength, given the intrinsic optical properties of the medium. The solution to that equation then becomes the forcing term of the second diffusion approximation, modeling the propagation of light at the fluorescence wavelength. The direct solution of this coupled problem is complicated [21], and has led to the development of normalization methods for use when only the fluorescence distribution is of interest [4, 22, 23]. To compute inverse solutions, the underlying physical models are typically linearized, to yield a linear algebraic form for the inverse problem. The forward matrices involved are highly ill-posed, and require regularization to stabilize solution of the inverse problem.

1.2 Thesis Contributions

This thesis makes three primary contributions to the advancement of FMT imaging techniques. The first consists of the derivation and analysis of an approximate statistical model for normalized fluorescence imaging. This results in two methods for obtaining approximate maximum likelihood solutions to the problem of fluorescence molecular tomography using the Born ratio. The second examines methods for utilizing CT or other

structural information as *a priori* information within the FMT reconstruction. The final contribution offers approaches to compensate for errors within structural prior information. These contributions are completed within the context of *in vivo* imaging of several mouse disease models, including breast cancer, lung inflammation and Alzheimer’s disease.

Statistical Inversion For imaging situations where the spatial distribution of fluorochrome is the only desired image, a data normalization scheme referred to as the normalized Born ratio has previously been introduced [4, 22, 23]. By using data collected at the excitation wavelength to normalize the corresponding fluorescence measurements, the Born ratio allows for direct determination of fluorescence distributions without explicitly solving the full coupled set of differential equations. It achieves this by providing a degree of invariance to inhomogeneities in the background optical parameters of the medium [23]. Additionally, use of the Born ratio eliminates the need to determine source and detector coupling coefficients, further simplifying the inversion process.

While the Born ratio is useful for overcoming nonlinearities arising from imaging through an inhomogeneous medium, it also introduces a problem. Typically in optical imaging, a higher amplitude received signal is associated with a higher signal to noise ratio (SNR) (i.e., more information content). However, when using the Born ratio, the processed data are the ratio of two such signals, and thus the absolute intensity is no longer an indicator of SNR. This has led to data preprocessing methods which involve thresholding, using only those data points whose component signals are judged to have sufficiently high SNRs. While effective, there exists no analytic method for setting these thresholds, leaving these methods highly subjective and operator dependent. This dependency leads to increased variability between individual inversions and is highly undesirable.

To eliminate this variability, the first contribution made by this thesis is the derivation of a statistically-based inversion method which avoids the use of these subjective thresholds. Beginning with statistical analysis and model determination for the received

signals, we determined an appropriate statistical model for the Born ratio. Several approximations were then made and justified to arrive at a simplified form of the probability density. From this, an approximate maximum likelihood solution was derived in the form of a fixed point iteration [24]. Finally, a further reduced solution was described, implemented as a diagonally weighted least squares problem.

MultiModality Imaging While the data collected by fluorescence based diffusion imaging systems contains large amounts of functional information, they are hindered by a lack of spatial resolution. One method which has been suggested to help improve on this limitation is the use of structural prior information, obtained from other imaging modalities such as X-ray computed tomography (CT) or magnetic resonance imaging (MRI) [25–29]. This information can then be incorporated into the tomography problem as an anatomically based prior model of the image being reconstructed [30, 31].

The development of appropriate methods for the inclusion of CT or MRI structural information in the construction of FMT inverse solutions constitutes the second primary contribution of this thesis. Our approach begins with a high resolution X-ray CT structural image, segmented into experimentally relevant anatomical regions. This segmentation leads to a partial volume labeling scheme which accurately represents the anatomic regions contributing to each solution voxel. From this, a reduced dimensional problem was constructed to extract information from the collected data regarding the likelihood of fluorescence within each region.

The construction and validation of this algorithm took place within the context of an *in vivo* study of the development of amyloid- β plaques in a mouse model of Alzheimer’s disease. The described method used the information from the low dimensional solution to construct a diagonal regularization term where the degree of regularization applied to each voxel was individualized based on the tissue(s) within which that voxel lay. This allowed voxels to update individually, and avoids imposing constraints such as smoothness on the relationships between voxels. While applied here to the study of Alzheimer’s

disease in mice, the method is general enough that it could easily be applied to other *in vivo* imaging targets with only the structural details being altered.

The final chapter of this thesis addresses two approaches designed to help compensate for errors in the structural prior information. These errors can arise either as a result of poor segmentation or registration between imaging modalities, or from fundamental properties of the underlying fluorescence. In particular, because the fluorochromes imaged by FMT target biochemical level activity, their physical distribution may not align precisely with anatomy. Physically apparent disease progression may be preceded by changes on a biochemical level. The ability to accurately image these changes *in vivo* would be a valuable tool for biomedical imaging.

The first approach to correcting such errors uses differential equations as the basis for a more complex Gaussian image prior. Through the incorporation of boundary conditions between each anatomic region, correlation is established between neighboring regions. This allows information to be passed across boundaries, and thus also allows for the more accurate reconstruction of fluorescent distributions whose spatial distribution does not align perfectly with the anatomic structure. The use of differential equations to define the prior model provides a natural framework for defining maximum a posteriori (MAP) estimation problems, and allows for a wide range of behaviors to be accurately modeled.

As an alternate approach to allowing a “floating” boundary location, we describe a technique that uses the sparsity inducing properties of 1-norm minimization. This method begins by defining a boundary region that extends for some distance from the nominal *a priori* boundary location. In the “trusted” regions, which are not part of the boundary region, traditional 2-norm based smoothing regularization is applied. Within the boundary region, vectors normal to the boundary are determined, and the image gradient along them computed. The 1-norm of this gradient is then used as a constraint in the final optimization problem. By using a convex optimization routine, which minimizes

this 1-norm, subject to constraints on data and regularization mismatch, a boundary is obtained, which can move freely within the defined boundary region, as determined by the data itself.

1.3 Optical Medical Imaging

The use of light to extract information about biological tissues is not a new concept. Microscopy has been around for centuries, and examines the light reflected from at, or just below, the surface of a medium [32]. A myriad of advances have been made over the centuries, making microscopy in its multiple current forms among the most commonly used imaging techniques in science.

The desire to see deeper into tissue has, however, spurred the development of more complex techniques. Initial efforts can be dated as far back as the early twentieth century, with the suggestion to image breast lesions with transillumination [33]. Diffuse light has also found non-imaging applications, such as in the ubiquitous pulse oximetry systems [34]. However significant progress for tomographic purposes was not made until recent years, when advances in both optical equipment, mathematical models, and computing power have made accurate diffusion imaging possible.

1.3.1 Diffuse Optical Tomography

One of these new methods, commonly referred to as Diffuse Optical Tomography (DOT), uses light in the near-infrared range to determine spatial maps of the intrinsic optical absorption and scattering coefficients present within tissue [35–37]. This range of wavelengths, between approximately 650nm and 900nm, corresponds to a gap in the absorption spectrum of water, allowing imaging of other biologically relevant chromophores such as hemoglobin and deoxyhemoglobin [38, 39]. These molecules, which are closely associated with blood volume and oxygen saturation, enable imaging applications such as breast cancer detection, functional brain imaging, and fetal oxygen monitoring [40–42].

Furthermore, by making use of relationships between physical chromophore concentrations and the resulting optical parameters, multispectral methods have been implemented which can directly determine the concentrations of target chromophores without intermediate solution of the absorption and scattering parameters [43–47].

1.3.2 Radiative Transport

Due to the diffusive behavior of infrared radiation in tissue, deterministic modeling methods based on Maxwell’s equations have such huge computational requirements associated with them that they are impossible to implement. Because of this, methods based on stochastic relations have been developed to deal with these types of problems. Given the enormous number of photons involved in a typical experiment, it is useful to model them in a bulk stochastic fashion, rather than treating each individually. The resulting “radiance” or “photon density” can be thought of as being related to the number of photons in a given volume. The equation forming the basis of these methods is the Radiative Transport Equation (RTE), originally developed by Chandrasekhar [48]:

$$\frac{1}{v} \frac{\partial L(\mathbf{r}, \hat{\Omega}, t)}{\partial t} + \nabla \cdot L(\mathbf{r}, \hat{\Omega}, t) + \mu_t L(\mathbf{r}, \hat{\Omega}, t) = \mu_s \int L(\mathbf{r}, \hat{\Omega}', t) f(\hat{\Omega}, \hat{\Omega}') d\hat{\Omega}' + S(\mathbf{r}, \hat{\Omega}, t) \quad (1.1)$$

In (1.1), $L(\mathbf{r}, \Omega, t)$ is the radiance (a measure of the relative concentration of photons) at a position \mathbf{r} , propagating along the spherical angle Ω , at time t . The probability of a photon scattering from direction $\hat{\Omega}$ to direction $\hat{\Omega}'$ is given by $f(\hat{\Omega}, \hat{\Omega}')$. The speed of light is c and the source is denoted as $S(\mathbf{r}, \hat{\Omega}, t)$. Additionally, μ_a and μ_s are the spatially varying absorption and scattering parameters, respectively, with μ_t equal to $\mu_s + \mu_a$. The terms on the left side represent radiance lost from a differential volume in a differential solid angle, while the terms on the right represent the radiance gained. Thus this equation is an expression of conservation of photons.

1.3.3 Diffusion Approximation

While inversion methods making use of the RTE have emerged in recent years [8, 49–51], the amount of computation involved is still typically more than can be applied for general inverse problems with large data sets. As such, given source-detector separations greater than a few millimeters, and several assumptions about the properties of the source and diffusive material, the diffusion approximation is typically used in place of the RTE. Assuming an isotropic source, and $f(\hat{\Omega}, \hat{\Omega}')$ which is dependent only upon the angle between directions, the diffusion approximation can be derived [39, 52], which states that the photon density, $\Phi(r)$, satisfies the differential equation:

$$(\mathbf{D}\nabla^2 - \mu_a)\Phi(\mathbf{r}) = -s(\mathbf{r}) \quad (1.2)$$

$$\mathbf{D} = \frac{1}{3[\mu_a + (1 - \bar{p})\mu_s]}$$

where v is the speed of light in the tissue, μ_a and μ_s are the absorption and scattering coefficients, respectively and $s(\mathbf{r})$ is the source term. The parameter \bar{p} represents the mean cosine of the scattering angle and, given the assumption of isotropic diffusion, replaces the angular dependence seen in the radiative transport equation.

For certain specific homogeneous geometries such as an infinite volume, infinite halfspace, and uniform thickness slab, analytic solutions to the diffusion approximation can be obtained in the form of Green's functions [39]. These functions are solutions to the diffusion approximation when the source term is a spatial delta function. The solutions to two such geometries will be of use in this thesis: the infinite space and slab geometries. The infinite space geometry assumes that the diffusing medium has no boundaries, or that the boundaries are sufficiently distant that they will have a negligible effect upon the solution. In this case, the solution to (1.2), given $s(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{r}_s)$ will be [53]:

$$G(\mathbf{r}_s, \mathbf{r}) = \frac{\exp\{i\mathbf{D}|\mathbf{r} - \mathbf{r}_s|\}}{|\mathbf{r} - \mathbf{r}_s|} \quad (1.3)$$

That is, given a source located at \mathbf{r}_s , and a distance $|\mathbf{r} - \mathbf{r}_s|$ between that point and a point \mathbf{r} , $G(\mathbf{r}_s, \mathbf{r})$ is the expected photon density at location \mathbf{r} .

To construct solutions for geometries involving air-tissue interfaces, the boundary conditions at those interfaces must be established and satisfied. While multiple potential boundary conditions exist [54], one commonly used condition is the zero-boundary or extrapolated boundary condition. This condition states that the photon density should take a value of zero either along the physical boundary, or at all points a specified distance outside the boundary. To satisfy this condition for the slab geometry, an infinite summation of positive and negative infinite space Greens functions (1.3) is used to force the solution to zero at the desired boundary location.

More complex geometries require correspondingly more complex methods for solution. For arbitrary geometries, rather than attempt to obtain analytic solutions, the most common methods instead solve the diffusion approximation numerically. Arising in such forms as the finite difference (FD), boundary element (BE), and finite element (FEM) methods, these numerical solutions are used in place of the analytical Green's functions for computation of both forward and inverse solutions. Other solutions have been offered which attempt to find analytical solutions for arbitrary geometries using discretized surfaces, plane waves, and other mathematical methods [55–57].

1.3.4 Born Approximation

While the forward problem (that is, determining the value of $\Phi(\mathbf{r})$, given $s(\mathbf{r})$) has a closed form given assumptions about homogeneity and geometry, and can be solved in the general case using various numerical methods, the inverse problem is significantly more complicated. Direct inverse solutions to the diffusion equation are very difficult, and as such, the problem is frequently discretized and linearized in some manner to allow solutions to be more easily obtained. One commonly used method, arising from perturbation theory, is the Born approximation. This model assumes that the total received signal is equal to the expected signal given a known background, plus a perturbation due to differences between the background and the actual optical parameters. This yields the

model [58]:

$$\mathbf{u}_{total} \simeq \mathbf{u}_{background} + \Delta\mathbf{u}(\mathbf{x}) \quad (1.4)$$

where \mathbf{x} is the perturbation in fluorochrome concentration or optical properties from the assumed background. In practice, using traditional DOT systems, the value of $\mathbf{u}_{background}$ is simply computed and subtracted from the received data to obtain the perturbation data. For fluorescence based systems, $\mathbf{u}_{background}$ is typically assumed to be approximately zero [22]. For experiments with non-negligible background fluorescence, analytic expressions have been developed which allow for the subtraction of this background signal, allowing for accurate imaging even in its presence [23].

Under the Born approximation, the computation of $\Delta\mathbf{u}(\mathbf{x})$ assumes that the effects of a perturbation at a particular point within the volume are unaffected by perturbations at other locations, and that the total perturbation is simply the sum of all individual first order scattering events [36]. This results in:

$$\Delta\mathbf{u}(\mathbf{r}_s, \mathbf{r}_d) \simeq \int_{\Omega} d^3\mathbf{r} G(\mathbf{r}_d, \mathbf{r}) \mathbf{x}(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_s). \quad (1.5)$$

Here, \mathbf{r}_s and \mathbf{r}_d are the source and detector locations, respectively. The Green's function $G(\mathbf{r}, \mathbf{r}_s)$ describes the propagation of light from the source, through a homogeneous medium to the point \mathbf{r} . The perturbation $\mathbf{x}(\mathbf{r})$ interacts with the light, and is then scattered to the detector by the Green's function $G(\mathbf{r}_d, \mathbf{r})$. Integrating across the entire volume gives the total response seen due to first order scattering events.

1.3.5 Linearization and Formation of Inverse Problem

Because computing an inverse solution directly from the above integral equation is difficult, the spatial domain is commonly discretized, converting the integral into a summation across a discrete basis. The most commonly chosen is a piecewise constant voxel basis. This reduces the above equation to:

$$\Delta\mathbf{u}(\mathbf{r}_s, \mathbf{r}_d) = \sum_{\mathbf{r}_j \in \Omega} G(\mathbf{r}_d, \mathbf{r}_j) \mathbf{x}(\mathbf{r}_j) G(\mathbf{r}_j, \mathbf{r}_s) \Delta V_j \quad (1.6)$$

which represents the value of a single data point as a linear combination of the image values at each of the voxel locations \mathbf{r}_j . Because the value is constant across the voxel, the integral is replaced by simple multiplication with the voxel volume ΔV_j , and \mathbf{r}_j is typically taken to be at the center of the voxel. Using a vector-vector product to represent the summation, an equation for the i^{th} source-detector pair can be rewritten as:

$$\Delta \mathbf{u}(\mathbf{r}_{si}, \mathbf{r}_{di}) = \mathbf{w}_i \mathbf{x}, \quad (1.7)$$

where the j^{th} elements of \mathbf{w}_i and \mathbf{x} are $G(\mathbf{r}_{di}, \mathbf{r}_j)G(\mathbf{r}_j, \mathbf{r}_{si})$ and $\mathbf{x}(\mathbf{r}_j)$, respectively. When all source-detector pairs are taken into consideration, the resulting forward problem can be written as the linear system:

$$\mathbf{u} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_M \end{bmatrix} \mathbf{x} = \mathbf{W} \mathbf{x} \quad (1.8)$$

where M is the total number of source-detectors pairs. We have eliminated Δ from the equation for simplicity, and will continue to do so for the remainder of this thesis. This equation can be used to take a known distribution of optical parameters, \mathbf{x} , and generate a vector of expected data values \mathbf{u} . The inverse problem, however, is to determine the value of \mathbf{x} , given a known \mathbf{u} . For reasons discussed below in §2.2, an estimate of \mathbf{x} can be obtained by solving the Tikhonov regularized least squares problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{u} - \mathbf{W} \mathbf{x}\|_2^2 + \lambda^2 \|\mathbf{L} \mathbf{x}\|_2^2. \quad (1.9)$$

While analytic solutions to this problem are readily available, they are often computationally expensive. Instead, solution of the above problem is typically solved using numerical methods such as the conjugate gradient (CG) method, LSQR, and others. Further discussion of ill-posed problems and regularization, as well as details of several inversion algorithms, are in Sections 2.1 and 2.3.

1.4 Fluorescence Diffusion Imaging

Given the proliferation of fluorescent probes available for use in general microscopy, the extension of diffuse optical methods to systems involving fluorescing targets seems natural. An excitation laser is chosen with a wavelength such that it will sufficiently excite the fluorochrome, and filters are used to select the appropriate wavelength of light exiting the medium. Because the collected data are affected by the intrinsic optical properties as well as the presence of fluorochromes, data are generally collected at both the excitation and emission wavelengths. Data at the excitation wavelength contain information about the intrinsic properties of the medium, while the emission wavelength data additionally incorporate the presence of fluorochromes.

The resulting physical model is governed by not one, but two, coupled differential equations. The first of these describes the diffusion of light at the wavelength of the excitation laser. The second uses the diffuse photon field defined by the first equation as the forcing term to describe the diffusion of light at the fluorescent wavelength. These equations, which are dependent upon both the optical properties of the tissue and the frequency at which the intensity of the source laser is modulated, can be written as [13,21]:

$$-\nabla \mathbf{D}_x(r) \nabla \Phi_x(r, \omega) + \left(\mu_{a_x}(r) + \frac{i\omega}{c(r)} \right) \Phi_x(r, \omega) = q_0(r, \omega) \quad (1.10)$$

$$-\nabla \mathbf{D}_m(r) \nabla \Phi_m(r, \omega) + \left(\mu_{a_m}(r) + \frac{i\omega}{c(r)} \right) \Phi_m(r, \omega) = \Phi_x(r, \omega) \nu \mu_{a_f}(r) \frac{1 - i\omega\tau(r)}{1 + [\omega\tau(r)]^2} \quad (1.11)$$

Where $q_0(\mathbf{r}, \omega)$ is the source term, $\mu_{a_x}(\mathbf{r})$ and $\mathbf{D}_x(r)x(\mathbf{r})$ are the absorption and diffusion coefficients, respectively, and $\Phi_x(\mathbf{r}, \omega)$ is the photon density at the excitation wavelength. The modulation frequency of the source laser intensity (if any) is denoted by ω and $c(\mathbf{r})$ is the speed of light in the medium. At the emission wavelength, $\mathbf{D}_m(r)x(\mathbf{r})$ and $\mu_{a_m}(\mathbf{r})$ are the optical parameters, while $\Phi_m(\mathbf{r}, \omega)$ is the photon density. It should be noted that the continuous wave case can easily be derived from the above frequency domain equation by simply setting $\omega = 0$.

One method of solution, undertaken by several research groups, directly solves the coupled system in a two step process [13,14,21,59]. Using data collected at the excitation wavelength, **a solution to the first equation is obtained**. This yields a spatial map of the absorption and scattering parameters, allowing for evaluation of the excitation photon field within the volume. Given this information, a solution to the second equation can be computed, to yield the spatial distribution of fluorescence. This method introduces additional complexity by requiring determination of the background optical properties of the medium, parameters which may not be of interest for fluorescence molecular imaging.

1.4.1 Normalized Born Ratio

As an alternative to directly solving the joint problem, a method known as the normalized Born ratio, or simply the Born ratio, was introduced in [4, 22, 23]. This normalization scheme seeks to simplify the process of obtaining a solution when the only parameter of interest is the distribution of fluorochrome within the tissue. By using data at the excitation wavelength to normalize the collected fluorescence data, direct solution for the fluorochrome distribution is made possible. Assuming a homogeneous medium, the excitation wavelength Green's function $G_o^{\lambda_{exc}}(\mathbf{r}_d, \mathbf{r}_s)$ can be used to write an equation for the excitation measurements $U_{exc}(\mathbf{r}_d, \mathbf{r}_s)$ as:

$$U_{exc}(\mathbf{r}_d, \mathbf{r}_s) = Q_E^{\lambda_{exc}} \Theta_f^{exc} \Theta_{det}(\mathbf{r}_d) G_o^{\lambda_{exc}}(\mathbf{r}_d, \mathbf{r}_s) \Theta_{src}(\mathbf{r}_s) \quad (1.12)$$

with quantum efficiency $Q_E^{\lambda_{exc}}$, excitation filter attenuation Θ_f^{exc} . The source and detector coupling coefficients, which measure the efficiency with which photons are transferred from the laser into the tissue, and then from the tissue to the CCD detector, are denoted $\Theta_{src}(\mathbf{r}_s)$ and $\Theta_{det}(\mathbf{r}_d)$, respectively. Given fluorescent measurements modeled as:

$$U_{fl}(\mathbf{r}_d, \mathbf{r}_s) = Q_E^{\lambda_{fl}} \Theta_f^{fl} \int d^3r \Theta_{det}(\mathbf{r}_d) G_o^{\lambda_{fl}}(\mathbf{r}_d, \mathbf{r}) \frac{\nu}{D^{\lambda_{fl}}} x(\mathbf{r}) G_o^{\lambda_{exc}}(\mathbf{r}, \mathbf{r}_s) \Theta_{src}(\mathbf{r}_s) \quad (1.13)$$

where $Q_E^{\lambda_{fl}}$ is the quantum efficiency of the detector at the fluorescence wavelength, Θ_f^{fl} is the attenuation of the fluorescence filter, while ν is the speed of light within the

medium and $D^{\lambda_{fl}}$ is the diffusion coefficient at the fluorescence wavelength. The parameter $x(r)$ represents the spatially varying fluorochrome concentration that we want to recover. Finally, $G_o^{\lambda_{exc}}(r, \mathbf{r}_s)$ and $G_o^{\lambda_{fl}}(\mathbf{r}_d, r)$ are the two point Green's functions describing the transport of light from source to voxel and voxel to detector at the corresponding wavelengths.

Given these expressions, the so-called Born ratio relationship can be written as:

$$\begin{aligned} U(\mathbf{r}_s, \mathbf{r}_s) &\triangleq \frac{U_{fl}(\mathbf{r}_d, \mathbf{r}_s)}{U_{exc}(\mathbf{r}_d, \mathbf{r}_s)} \\ &= \alpha \frac{\int d^3r G_o^{\lambda_{fl}}(\mathbf{r}_d, r) x(r) G_o^{\lambda_{exc}}(r, \mathbf{r}_s)}{G_o^{\lambda_{exc}}(\mathbf{r}_d, \mathbf{r}_s)} \end{aligned} \quad (1.14)$$

with:

$$\alpha = \frac{Q_E^{\lambda_{fl}}}{Q_E^{\lambda_{exc}}} \frac{\Theta_f^{fl}}{\Theta_f^{exc}} \frac{\nu}{D^{\lambda_{fl}}} \quad (1.15)$$

The parameter α relies only upon basic system characteristics, and can be easily determined experimentally from measurements of a fluorochrome of known concentration [23]. Thus, once a simple system calibration has been done, there exist no excess free parameters to be determined.

1.5 Non-contact Imaging

Most existing optical diffusion systems are based around fixed imaging geometries. Imaging of endogenous contrast requires knowledge of source and detector coupling coefficients, and these can be made largely constant by coupling every optical fiber into an identical optical scattering liquid, rather than directly into tissue. The use of these so called matching fluids led to fixed imaging geometries such as a slab or cylinder. In addition to reducing the variability in coupling coefficients, analytic Green's function solutions to the diffusion equation are available for these geometries. Sources and detectors are localized using fixed optical fibers, and the geometry enforced by filling the imaging chamber with an optical matching fluid. When imaging fluorescence, however,

data are collected at multiple wavelengths, and often preprocessed to normalize the measurements [10, 55–57, 59–61]. This process of normalization eliminates or significantly reduces the problem of variable source and detector coupling coefficients, making the use of optical matching fluid unnecessary.

To use a fixed geometry, one of two conditions must be satisfied. Either all objects to be imaged must be the same shape, or they must be able to be placed within a closed imaging chamber of fixed geometry. Outside of controlled phantom studies, the first condition is unlikely to be satisfied, and thus the second must apply. To ensure that the entire chamber is diffusing, any volume not occupied by the imaging target is filled with a matching fluid, typically composed of a mixture of water, lipid solution, and India ink [4]. Varying the concentration of lipid and India ink allows for changes to be made in the scattering and absorption parameters, respectively. However, by using a matching fluid, light must propagate through additional diffusing media between the source and detector locations. This reduces the quality of the collected data below its theoretical maximum, and consequently degrades the final solution.

Additionally, many imaging systems are based on the use of optical fibers for illumination and data collection. While this method had been shown to be experimentally viable, it introduces problems of limited sampling density and fixed geometry. Reconfiguring optical fibers into an arbitrary geometry is generally not feasible on an experiment to experiment basis. Thus, the source and detectors configurations tend to be fixed. Furthermore, each fiber requires a finite amount of space, meaning that only a fixed number can be positioned on the surface of the target. Given that dense sampling arrangements have been shown to yield improved reconstruction images, the use of dense sampling grids provided by CCD camera based data collection methodologies offer further improvements over fiber based approaches [62].

By eliminating both imaging chamber and optical fibers, and instead focusing the laser source directly on the surface of the target with emission data collected via a CCD

camera, much higher spatial sampling densities can be obtained. Because its position is defined by the movement of mirrors or translation stages, an infinite number of source patterns can be defined. This allows for the optimum illumination pattern to be used for each individual experiment, rather than using a single, non-optimal configuration. Detectors, defined as small regions in the CCD image, can likewise be positioned as needed. By additionally incorporating data collection from multiple projection angles, more complete information about the diffusive nature of the imaging target can be gathered, further improving reconstruction quality [63].

1.5.1 Surface Reconstruction

One drawback of such non-contact systems is that accurate modeling of light diffusion within an arbitrary volume requires that the structure of the volume's surface be known. To do this, we employ a previously developed method which uses a volume carving algorithm to localize the surface [63–65].

After the fluorescence and excitation data, an additional set of images are collected using a photoluminescent plate placed such that the target is between the plate and the camera. The resulting image on the CCD is a silhouette of the target. By collecting these silhouettes at a large number of angles, information about the structure of the surface is gained. For each angle, the silhouette gives the volume carving algorithm information about which points in the reconstruction space are inside and outside of the volume. By compiling all of this information together, a full picture of the surface emerges, and can be defined as a triangulated surface such as in Figure 1.2.

1.6 Fluorescence Molecular Imaging Systems

A number of different fluorescence tomography systems have been developed over the years. Two such systems are used in this thesis, both developed in the Lab for Biooptics and Molecular Imaging (LBMI) at Massachusetts General Hospital. The first

of these is a second generation system, based on a slab geometry, and using a fixed, fiber-based source pattern. This system was used for validation of the statistical inversion method detailed in Section 3.1, and has been used for a number of different *in vivo* studies. The second system is a 360° imaging system, using a rotational mounting system to obtain diffusion data from multiple projections. Additionally, this system is a non-contact free-space system, and operates by focusing the source laser and detector CCD array directly on the surface of the animal.

1.6.1 Planar FMT System

The slab based imaging system used for some of the studies in this thesis was based on a 1.3cm thick slab geometry. A diagram of the system is shown in Fig 1.1. Illumination was achieved with a laser diode at 672nm, passed through an optical switch (DiCon Fiberoptics, Richmond, CA) to select from an array of 46 source fibers arranged on a $1.8 \times 1.2 \text{ cm}^2$ area. Output power at the imaging chamber was approximately 10mW. For detection, a CCD camera (Roper Scientific, Trenton NJ) electrically cooled to -35°C was used. Selection of excitation and emission wavelengths was achieved by appropriate filters : bandpass (three-cavity interference at $670 \pm 5 \text{ nm}$ for excitation measurements and $710 \pm 10 \text{ nm}$ for emission measurements (Andover, Salem NH)) and long-pass (emission wavelength: cutoff 695nm; Omega Optical, Brattleboro VT).

To collect data using this system, the imaging target was placed within the chamber, which was then compressed to a thickness of 1.3cm. The remainder of the chamber was then filled with a mixture of Intralipid and India ink, the concentrations of which were chosen such that the overall absorption and scattering parameters are similar to those seen in typical bulk tissue. For the so-called $1\times$ matching fluid, this required 60ml of Intralipid and 6ml of 1% India ink solution in a total solution volume of 600ml. Other matching fluid concentrations, denoted $1.5\times$, $2\times$, etc, were created by holding the amount of Intralipid in the solution constant and linearly varying the amount of India ink. Note

that Intralipid is both the brand name of the commercially available lipid solution used in preparation of the matching fluid, as well as the term used to refer to the final matching fluid itself. Where not otherwise specified, the term refers to the final, mixed, matching fluid.

1.6.2 Rotational Imaging System

The second system used was a 4th generation fluorescence molecular tomography system based on the multi-angle collection of transmission data in a freespace geometry. Through a combination of rotation and translation stages, a single fiber source is scanned across the surface of the object and data is collected via a CCD camera in a transmission geometry. This is repeated for multiple angles, yielding full view diffusive data. A diagram of the system is shown in Fig. 1.2.

A single optical fiber is used to channel light from a 650nm diode laser to illuminate the target. The free end of the fiber is mounted on a set of two translation stages (Velmex, Bloomfield NY), oriented orthogonal to the source-camera axis, and yielding a 2.5cm range of motion along each of the X and Y axes. This configuration allows for the source to be localized to any point on the surface facing the stages, resulting in the transmission geometry previously mentioned. Filter mounts placed between the fiber aperture and imaging target allow for the placement of focusing and/or attenuation elements, as required by the experiment. Fluorescence images used only the focusing lenses to provide maximum power to the target, while excitation images used neutral density filters to prevent saturation of the CCD camera.

The object to be imaged is loaded onto a centrally located rotation stage (Velmex, Bloomfield NY) via one of several interchangeable mounting inserts. Mounting for a phantom will be different than for imaging the torso of a mouse, which will be different than mounting for brain imaging. This system of multiple inserts allows for the optimal method to be chosen on a per-experiment basis, and altered as necessary.

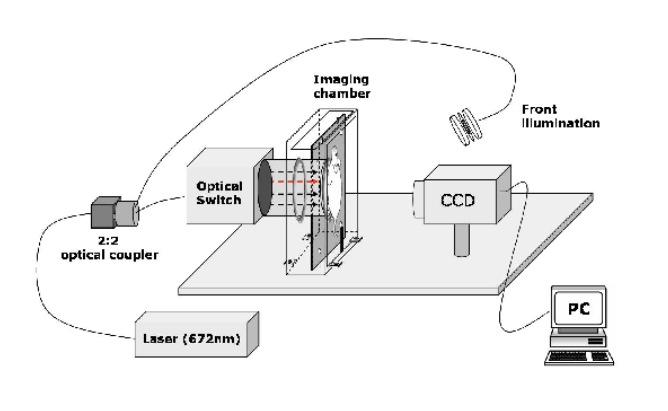


Figure 1.1: **Diagram of Planar FMT Imaging System** Laser light from the source laser is coupled into the imaging volume through an optical switch and 46 optical fibers arranged in a grid pattern at the back of the imaging chamber. Data collected is done in parallel using a CCD camera and virtual detectors defined as small integrated regions of the CCD.

Measurements are collected using an electrically cooled CCD (Roper Scientific, Trenton NJ) positioned on the opposite side of the chamber from the laser fiber. A multi-position filter wheel is used to select wavelength appropriate filters. For the experiments detailed herein, the excitation measurement was collected using a bandpass filter, while fluorescence measurements were collected with a combination of a bandpass (three-cavity interference at 710 ± 10 nm, Andover, Salem MA) and a longpass filter (Cutoff 695nm; Omega Optical, Brattleboro VT). Typical data collection configurations collect data from approximately 18 rotation angles.

Synchronization of the above components is implemented using commercially available controller hardware (Velmex, Bloomfield NY), and a custom Matlab GUI interface. Setup consists of defining the source location grid and specifying the angles at which data is to be collected. The system then scans the target, collecting one image for each source location in the grid, before progressing to the next rotation angle. This process is then repeated until data has been collected for all angles.

To obtain surface information for later use in modeling a further set of data is collected, replacing the laser source with an electroluminescent plate. Single frames of data are taken at each of an increased number of angles, typically sampling every degree. These images are then used in conjunction with a volume carving algorithm to reconstruct the surface of the object being imaged.

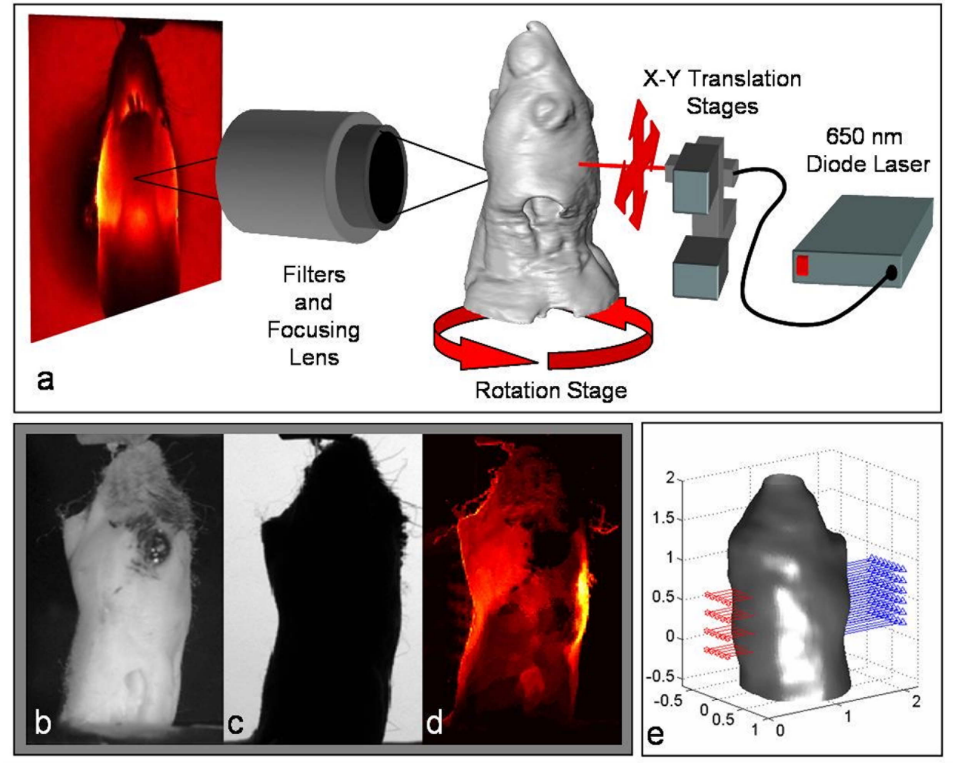


Figure 1.2: **Rotational FMT System and Localization of Animal Surface** a) Schematic of data collection system. Source fiber driven by a 650nm diode laser was mounted on translation stages for focusing and localization on animal surface. Collection of data was done in a transmission geometry using an electrically cooled CCD camera and appropriate filters for wavelength selection. A representative excitation channel image is shown. b) White light image of mouse head. c) Corresponding silhouette image used for surface localization. 360 such images were used for the surface generation algorithm. d) Summation of normalized (fluorescence/excitation) data across all source locations for this projection angle. e) Resolved surface location, showing source (red) and detector (blue) patterns mathematically projected onto surface for a single projection angle.

Chapter 2

Inverse Problems and Solution Methods

2.1 Introduction to Inverse Problems

There are many aspects of our world for which direct observation is either impossible, impractical, or undesirable for one reason or another. Astronomers seek to see into the far reaches of our universe, while geophysicists probe the depths of our own planet. In medicine, doctors want to visualize the internal workings of a patient, without resorting to the scalpel. And in some cases, such as divining the molecular scale activity of cancers and other diseases, direct observation can significantly disturb the activity being analyzed. In applications such as these, inverse solutions allow unobservable model parameters to be estimated based on measurements of physical processes which have interacted with that model. This is in contrast to the so-called forward problem, where the mathematical model is used in conjunction with a known set of parameters to generate a set of expected ~~data~~ measurements.

For medical imaging applications, a wide range of physical processes are used to probe the tissue. Ultrasound, electromagnetic waves, magnetic fields, electric voltage potentials, X-rays, and other physical phenomena have all been exploited to measure the internal characteristics of tissue [66, 67]. Each of these applications is associated with a mathematical model which describes the physical process underlying the data collection, and relates the data to physical parameters of interest. Some of these physical models can be analytically "inverted" to give an expression for the physical parameters in terms of

the measured data. This is not generally the case for fluorescence molecular tomography and related diffusion based imaging techniques, although some analytic approaches have been proposed for certain special cases [68,69]. Instead, diffusion based imaging systems result in what are referred to as “ill-posed” inverse problems that must be solved by stabilized numerical approaches.

For a large class of inverse problems, the collected data are assumed to be linearly related to the model parameters, either globally, or in a small region of the solution space about some known point. This assumption leads to a linear equation representing the physical system:

$$\mathbf{W}\mathbf{x} \simeq \mathbf{u} = \mathbf{u}_{true} + \mathbf{n} \quad (2.1)$$

Here, \mathbf{x} is the vector of model coefficients to be recovered, and \mathbf{W} is a matrix describing the physical relationship between the model parameters and the collected data. The collected data are \mathbf{u} , comprised of two components. The first, \mathbf{u}_{true} , is the portion of the data which is consistent with the model (i.e. $\mathbf{W}\mathbf{x} = \mathbf{u}_{true}$ is satisfied for some \mathbf{x}). The second term, \mathbf{n} , is used to represent both measurement noise as well as errors resulting from mismatch between the linear model \mathbf{W} and the true physical process, which is nonlinear in the case of FMT as well as many other imaging problems.

In general, the matrix \mathbf{W} is not square, and so naive solution of (2.1) as:

$$\hat{\mathbf{x}} = \mathbf{W}^{-1}\mathbf{u} \quad (2.2)$$

is not possible, as the inverse does not exist. Depending on whether the problem is under or overdetermined (that is, whether there are more or fewer unknowns than there are data points), and whether the data \mathbf{u} is in the range of \mathbf{W} there will be either infinitely many solutions which satisfy the equation, or no solution will exist. To solve the inverse problem in these situations, (2.1) is replaced with an optimization problem which seeks to find the set of parameters for which the 2-norm difference between the collected and

expected data is minimized:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{u} - \mathbf{W}\mathbf{x}\|_2^2 \quad (2.3)$$

Solution of this problem leads to the classical least squares solution (also known as the Moore-Penrose generalized matrix inverse, or alternately, the pseudo-inverse) [70]. For overdetermined problems, this takes the form:

$$\hat{\mathbf{x}} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{u}. \quad (2.4)$$

For underdetermined solutions, an infinite number of valid solutions exist, with the minimum norm solution \mathbf{x} , subject to $\mathbf{W}\mathbf{x} = \mathbf{u}$, given as:

$$\hat{\mathbf{x}} = \mathbf{W}^T (\mathbf{W}\mathbf{W}^T)^{-1} \mathbf{u} \quad (2.5)$$

In both cases, these formulations assume that \mathbf{W} has full rank, which is in general untrue for ill-posed inverse problems. In these cases, solution by direct use of this equation (if explicit matrix inversion is possible) will result in highly unstable solutions. Being ill-posed, the ratio between the highest and lowest singular values is exceedingly high [71]. When computing a solution, random noise present in the data will be amplified by its correlation with less significant singular vectors, resulting in solutions which are heavily corrupted by noise. To counter this issue, solutions are often obtained using the modified cost function:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{u} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda^2 \|\mathbf{L}\mathbf{x}\|_2^2. \quad (2.6)$$

The second term in this equation is referred to as a Tikhonov regularization term, which serves to stabilize the solution by providing a prior model for the reconstructed image x [72].

The cost function above corresponds to maximum a posteriori (MAP) estimator given a particular selection of image and noise models as multivariate Gaussians. A multivariate Gaussian distribution takes the form:

$$f(x_1, x_2, \dots, x_N) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right). \quad (2.7)$$

Given a zero mean Gaussian noise model ($\mathbf{n} \sim N(0, \Sigma_n)$), and assuming that the $\mathbf{W}\mathbf{x} = \mathbf{u}_{true}$, this yields an overall signal model of $\mathbf{u} \sim N(\mathbf{W}\mathbf{x}, \Sigma_n)$. If this is then combined with a general multivariate Gaussian for the image ($\mathbf{x} \sim N(\mu_x, \Sigma_x)$), the maximum a posteriori estimate of \mathbf{x} can be written as:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{u}, \mathbf{x}). \quad (2.8)$$

The joint probability can then be rewritten, and the log-likelihood taken, to obtain:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \log(p(\mathbf{u}|\mathbf{x})) + \log(p(\mathbf{x})). \quad (2.9)$$

Looking back at (2.10), we can see that the log of a multivariate Gaussian will be:

$$\log(f(x_1, x_2, \dots, x_N)) = \log\left(\frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}\right) - \left(\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right). \quad (2.10)$$

For a fixed Σ , the first term will be a constant, and subsequently will not affect any optimization routines. Using the second term, in combination with (2.9) and the stochastic models for \mathbf{u} and \mathbf{x} , we obtain:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} -\frac{1}{2}(\mathbf{u} - \mathbf{W}\mathbf{x})^T \Sigma_n^{-1}(\mathbf{u} - \mathbf{W}\mathbf{x}) - \frac{1}{2}(\mathbf{x} - \mu_x)^T \Sigma_x^{-1}(\mathbf{x} - \mu_x). \quad (2.11)$$

By turning the maximization of two negative components into the minimization of two positive components, eliminating the constant $1/2$ terms, and using the identity: $(\mathbf{a} - \mathbf{b})^T \mathbf{C}(\mathbf{a} - \mathbf{b}) = \|\mathbf{C}^{1/2}(\mathbf{a} - \mathbf{b})\|_2^2$, we can write:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\Sigma_n^{-1/2}(\mathbf{u} - \mathbf{W}\mathbf{x})\|_2^2 + \|\Sigma_x^{-1/2}(\mathbf{x} - \mu_x)\|_2^2. \quad (2.12)$$

which looks very similar to the expression seen in (2.6). In fact, any Tikhonov regularized least squares problem can be interpreted as a MAP solution given a particular choice of image and noise models. The regularization parameter λ , serves to adjust the relative variances of the noise and image models, when exact values are unknown. Because multiplication by a constant will not change the minimization, the standard Tikhonov model can be interpreted as assuming white Gaussian noise, where the variance has been

absorbed into the λ term. Likewise, the standard $\|\mathbf{L}\mathbf{x}\|_2^2$ regularization term can be seen as making statements about the covariance structure of the image, while assuming that the image is zero mean. Specifically, the application of the regularization term $\lambda^2\|\mathbf{L}\mathbf{x}\|_2^2$ assumes a Gaussian image prior of $\mathbf{x} \sim N(\mathbf{0}, \frac{1}{\lambda^2}(\mathbf{L}^T\mathbf{L})^{-1})$.

While the analytic solution to the above problem is:

$$\hat{\mathbf{x}} = (\mathbf{W}^T \Sigma_x \mathbf{W} + \lambda^2 \mathbf{L}^T \mathbf{L})^{-1} (\mathbf{W}^T \mathbf{u} - \mathbf{L}^T \mathbf{L} \mu_x) \quad (2.13)$$

solutions are most frequently obtained using numerical methods such as the conjugate gradient (CG) method, LSQR, and others. It is to these solution methods that we now turn our attention. We will discuss regularization and its relationship to the singular value decomposition (SVD) in section 2.3.

2.2 Solution Methods

A wide range of algorithms have been developed to solve least squares problems, both with and without explicit regularization. For this work, we make use of four of these methods. The first is the algebraic reconstruction technique (ART), used for comparison because of its previously established utility [73]. Given the specific structure of the fixed point iteration solution derived in section 3.1, the conjugate gradient method [74] is particularly well suited as a method of solution. For the reduced form of our maximum likelihood solution, the LSQR algorithm [75] is used to efficiently compute solutions using Krylov subspaces. Finally, we use the modified residual norm steepest descent (MRNSD) algorithm [76] to obtain non-negatively constrained solutions in a variety of situations.

2.2.1 Algebraic Reconstruction Technique

The algebraic reconstruction technique (ART) is an inversion method, commonly used in the construction of CT images, which obtains a solution by repeated projection onto the rows of the matrix \mathbf{W} [77]. The ordering of the projections is frequently random, to help eliminate any bias that may arise from using the same ordering multiple times.

With a typical initialization setting value of \mathbf{x}_0 to be all zeros, the estimate is updating using the iteration:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \lambda \frac{\mathbf{u} - \mathbf{W}_i \mathbf{x}}{\|\mathbf{W}_i\|_2^2} \mathbf{W}_i \quad (2.14)$$

A full iteration of the algorithm uses this update equation once for each vector in the basis set. Thus the above equation represents a single sub-iteration, with $\mathbf{x}^{(n)}$ and $\mathbf{x}^{(n+1)}$ representing the the estimate before and after the update is performed. \mathbf{W}_i is the i th row of the forward matrix, selected in a quasi-random fashion as described below. Finally, \mathbf{u} is the collected Born ratio data, and λ is a relaxation parameter, typically set to 0.1 for our experiments. Given the linearization in our model and the presence of noise in the data, the relaxation parameter is included in order to help stabilize the reconstruction.

To minimize the effect of projection order on the reconstruction, the rows of the matrix are accessed in a quasi-random order [73]. A data collection angle is first selected at random, and then a random ordering is selected for projection onto the rows associated with that angle. After each individual projection, all negative image values are truncated to zero to ensure a non-negative solution. Once complete, the process is repeated for each remaining collection angle, with a random ordering of the angles as well. This constitutes one iteration, and is repeated as necessary to obtain a reconstruction.

ART and Diagonal Weighting Matrices Later, in Section 3.1.4, I will derive a statistical inversion technique which applies a diagonal weighting matrix to the least squares minimization problem. It should be noted that neither this approach, nor any other diagonal weighting matrix, can be used with the ART algorithm. Because ART examines the data one point at time, the addition of a weighting parameter is ultimately canceled out in the update equation. Rewriting (2.14) to expand the norms and include

a weighting factor α , we get:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \lambda \frac{(\alpha \mathbf{W} - \alpha \mathbf{W}_i)}{(\alpha \mathbf{W}_i)^T (\alpha \mathbf{W}_i)} \alpha \mathbf{W}_i \quad (2.15)$$

$$= \mathbf{x}^{(n)} + \lambda \frac{\alpha^2 (\mathbf{W} - \mathbf{W}_i)}{\alpha^2 \mathbf{W}_i^T \mathbf{W}_i} \mathbf{W}_i \quad (2.16)$$

$$= \mathbf{x}^{(n)} + \lambda \frac{(\mathbf{u} - \mathbf{W}_i)}{\mathbf{W}_i^T \mathbf{W}_i} \mathbf{W}_i \quad (2.17)$$

which is exactly the same as (2.14), meaning that the diagonal weighting will have no effect upon the solution.

An alternative method that may warrant further exploration is the use of the diagonal weighting parameters to modulate ART steplength, rather than as an explicit weighting matrix. Using the same α to denote the weighting parameter, the update equation could then be written as:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \lambda(\alpha) \frac{(\mathbf{u} - \mathbf{W}_i)}{(\mathbf{W}_i)^T (\mathbf{W}_i)} \mathbf{W}_i \quad (2.18)$$

for some function $\lambda(\alpha)$, as yet to be defined.

2.2.2 Conjugate Gradient

The conjugate gradient (CG) algorithm is used to solve the linear system

$$\mathbf{W}\mathbf{x} = \mathbf{u} \quad (2.19)$$

for the case where \mathbf{W} is a real, symmetric, positive definite matrix [74]. While many inverse problems exist for which this condition is not satisfied, the CG algorithm can always be applied to the normal equations:

$$\mathbf{W}^T \mathbf{W} \mathbf{x} = \mathbf{W}^T \mathbf{u}. \quad (2.20)$$

This allows for the use of the algorithm for all general linear inverse problems.

Conjugate gradient is an iterative algorithm, belonging to the family of Krylov subspace methods, and like steepest descent, it uses the gradient of the cost function

as a method for determining a search direction and distance for each individual update. Unlike steepest descent, however, CG does not directly use the gradient as the search direction [74]. Instead, CG uses ***W**-conjugate* directions for the search, thereby giving rise to the algorithm's name. By using conjugate directions, CG ensures that each subsequent search direction is orthogonal to all previous search directions. This prevents repeatedly searching along the same direction, a problem that arises when using steepest descent, which can lead to an increased number of required iterations.

The algorithm starts by initializing the residual r , the search direction p , and the iteration number k :

$$\mathbf{r}_0 = \mathbf{u} - \mathbf{W}\mathbf{x}_0 \quad \mathbf{p}_0 = \mathbf{r}_0 \quad k = 0.$$

For each iteration, the optimal distance along the search direction is computed, then the residual and image vectors are updated:

$$\begin{aligned} \alpha_k &= \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{W} \mathbf{p}_k} \\ \mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{p}_k \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k \mathbf{W} \mathbf{p}_k \end{aligned}$$

A test for convergence is applied to the residual, and if it has converged, the algorithm is terminated. If another iteration is to be performed, the search direction is then updated as:

$$\begin{aligned} \beta_k &= \frac{\mathbf{r}_{k+1}^T \mathbf{r}_{k+1}}{\mathbf{r}_k^T \mathbf{r}_k} \\ \mathbf{p}_{k+1} &= \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k. \end{aligned}$$

Finally, the iteration number k is incremented, and another iteration of the CG algorithm begins. Further details, along with an in-depth discussion of the relationship between steepest descent and the conjugate gradient method, can be found in [74].

2.2.3 LSQR

LSQR is an iterative inversion algorithm, motivated by the Lanczos tridiagonalization process, which efficiently obtains solutions to the least squares minimization problem [75]. Using two bidiagonalizations, one of which reduces the matrix to upper bidiagonal form, while the other reduces it to lower bidiagonal form, a parallel problem can be constructed in a reduced dimensional Krylov subspace. This lower dimensional “projected” problem can be solved using a QR factorization, applied as a series of orthogonal transformations, without ever explicitly storing the associated matrices. The solution to the reduced dimensional problem can then be converted into a solution to the full minimization problem through a simple linear relation. The algorithm can be summarized as follows [75]. First, all necessary variables are initialized as:

$$\begin{aligned} \beta \boldsymbol{\vartheta}_1 &= \mathbf{u} & \alpha_1 \boldsymbol{\nu}_1 &= \mathbf{W}^T \boldsymbol{\vartheta}_1 & \mathbf{a}_1 &= \boldsymbol{\nu}_1 & \mathbf{x}_0 &= 0 \\ \bar{\phi} &= \beta_1 & \bar{\rho} &= \alpha_1 \end{aligned} \quad (2.21)$$

For each iteration, the next step of the bidiagonalization is first performed:

$$\beta_{i+1} \boldsymbol{\vartheta}_{i+1} = \mathbf{W} \boldsymbol{\nu}_i - \alpha_i \boldsymbol{\vartheta}_i \quad (2.22)$$

$$\alpha_{i+1} \boldsymbol{\nu}_{i+1} = \mathbf{W}^T \boldsymbol{\vartheta}_{i+1} - \beta_{i+1} \boldsymbol{\nu}_{i+1}. \quad (2.23)$$

Then the next orthogonal transformation in the QR factorization is applied:

$$\begin{aligned} \rho_i &= \sqrt{\bar{\rho}_i + \beta_{i+1}^2} \\ c_i &= \frac{\bar{\rho}_i}{\rho_i} \\ s_i &= \frac{\beta_{i+1}}{\rho_i} \\ \theta_{i+1} &= s_i \alpha_{i+1} \\ \bar{\rho}_{i+1} &= -c_i \alpha_{i+1} \\ \phi_i &= c_i \bar{\phi}_i \\ \bar{\phi}_{i+1} &= s_i \bar{\phi}_i. \end{aligned} \quad (2.24)$$

The solution and basis vectors are then updated as:

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \frac{\phi_i}{\rho_i} \mathbf{a}_i \quad (2.25)$$

$$\mathbf{a}_{i+1} = \boldsymbol{\nu}_{i+1} - \frac{\theta_{i+1}}{\rho_i} \mathbf{a}_i \quad (2.26)$$

A test for convergence is then applied, and the process repeated as necessary until convergence is obtained. Further details of the algorithm can be found in [75].

2.2.4 Modified Residual Norm Steepest Descent

A final algorithm which was used in the course of this work was the modified residual norm steepest descent (MRNSD) algorithm, as introduced by James Nagy [76]. This algorithm is a variation on the standard steepest descent algorithm, which uses the gradient of the cost function as the search direction, and simply implements the added constraint that the resulting solution cannot have any negative components. This is done by parameterization of the image and a restriction on the steplength of the algorithm. Using a parameterized version of the least squares cost functional $\Phi(x) = \|\mathbf{u} - \mathbf{W}\mathbf{x}\|_2^2$, with $\mathbf{x} = e^z$, the gradient with respect to z is:

$$\nabla_z \Phi(\mathbf{x}) = \text{diag}(\mathbf{x}) \nabla_{\mathbf{x}} \Phi(\mathbf{x}) = \text{diag}(\mathbf{x}) \mathbf{W}^T (\mathbf{W}\mathbf{x} - \mathbf{u}) \quad (2.27)$$

This gradient is used as the search direction for an iterative solution of the form:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)} \quad (2.28)$$

with $\mathbf{p}^{(k)} = -\nabla_z \Phi(\mathbf{x}^{(k)})$. Step length is chosen as either the optimal step length:

$$\alpha_{\text{optimal}} = \frac{(\mathbf{W}\mathbf{x} - \mathbf{u})^T \mathbf{W} \mathbf{X} \mathbf{W}^T (\mathbf{W}\mathbf{x} - \mathbf{u})}{\|\mathbf{W} \mathbf{X} \mathbf{W}^T (\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2}, \quad (2.29)$$

(with $\mathbf{X} = \text{diag}(\mathbf{x})$) or the maximum step length which maintains the non-negativity of the solution, whichever is greater:

$$\alpha_{\text{max}} = \min_{p_i < 0} (-x_i / p_i) \quad (2.30)$$

The step length used for the update is then chosen as the lesser of $\alpha_{optimal}$ and α_{max} . The iteration is repeated as necessary until some convergence criterion has been satisfied, or a set number of iterations have been completed.

2.3 Regularization

When dealing with the numerical solution of generalized linear systems, one concept which is repeatedly encountered is that of regularization. Regularization serves, first, to help counter the ill-posedness of many such problems, and allow for stable solutions to be obtained from noisy data. Second, regularization can serve as a means to incorporate additional information about the solution that is not encoded in the forward operator. Here, I give a short overview of regularization and how it applies to linear systems, and introduce several regularization methods which will be used later in this thesis.

2.3.1 Ill-Posedness

The primary reason for using regularization in the numerical solution of linear inverse problems is as a means to stabilize the solution and provide a level of robustness to noise for cases where the operator is highly ill-posed. As previously stated, a discrete ill-posed problem is said to be one in which the condition number of the matrix (that is, the ratio of highest to lowest singular values) is exceedingly high. This suggests that some (or all) of the columns of the matrix are close to linearly dependent upon one another.

A useful tool for examining ill-posed problems is the singular value decomposition (SVD). This matrix decomposition expresses an arbitrary $M \times N$ matrix \mathbf{A} as the product of three matrices [70, 78]:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^n \mathbf{u}_i \sigma_i \mathbf{v}_i^T \quad (2.31)$$

with $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \in R^{m \times n}$ and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \in R^{n \times n}$ being orthonormal matrices: $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}_n$. The matrix $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ is diagonal with ordered non-negative elements $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. The values σ_i are referred

to as the singular values, while the associated vectors u_i and v_i are the left and right singular vectors, respectively. These vectors can be seen to arise as the eigenvectors of the matrices $\mathbf{A}^T \mathbf{A} = \mathbf{V} \Sigma^2 \mathbf{V}^T$ and $\mathbf{A} \mathbf{A}^T = \mathbf{U} \Sigma^2 \mathbf{U}^T$. For an $M \times N$ matrix \mathbf{A} , the number of non-zero singular values will be equal to $rank(\mathbf{A})$.

The SVD can be used to write matrix multiplications such as $\mathbf{A}\mathbf{x}$ as:

$$\mathbf{A}\mathbf{x} = \sum_{i=1}^n \sigma_i (\mathbf{v}_i^T \mathbf{x}) \mathbf{u}_i \quad (2.32)$$

A well known property of singular vectors is that as one progresses from the singular vectors associated with high singular values to those associated with low singular values, the degree of oscillatory behavior seen in the singular vectors increases [71]. When computing a matrix vector product as above, correlation with the more highly oscillatory singular vectors is damped by the correspondingly lower singular values. If, however, the SVD is used to write a pseudoinverse of A as:

$$\mathbf{A}^\dagger = \sum_{i=1}^{rank(\mathbf{A})} \mathbf{v}_i \sigma_i^{-1} \mathbf{u}_i^T \quad (2.33)$$

then those highly oscillatory singular vectors are amplified by the inverse of the associated low singular values, rather than damped. This can create a problem when computing numerical inversions. If the data are noise free and perfectly match the model (an unrealistic scenario outside of simulations), then a suitable inverse solution will be found. If, however, the system includes noise and model error, represented jointly as \mathbf{n} :

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{exact} + \mathbf{n} \quad (2.34)$$

then the least squares inverse solution will be:

$$\tilde{\mathbf{x}} = \sum_{i=1}^{rank(\mathbf{A})} \sigma_i^{-1} (\mathbf{u}_i^T \mathbf{b}) \mathbf{v}_i = \sum_{i=1}^{rank(\mathbf{A})} \sigma_i^{-1} (\mathbf{u}_i^T \mathbf{A} \mathbf{x}_{exact} + \mathbf{u}_i^T \mathbf{n}) \mathbf{v}_i. \quad (2.35)$$

Here, the random noise present in the signal will correlate with the oscillatory singular vectors, and result in a solution which is far from the desired value of \mathbf{x}_{exact} .

One established method to compensate for this deficiency is known as Tikhonov regularization, and involves the addition of a second term to the minimization equation. Thus $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ becomes:

$$\|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{Lx}\|_2^2 \quad (2.36)$$

where \mathbf{L} is an arbitrary matrix (often simply the identity), and λ is a parameter controlling the degree of regularization. For the case where $\mathbf{L} = \mathbf{I}_n$, the SVD of \mathbf{A} can be used to analyze the least squares problem. Minimization of the above cost functional yields a solution:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I}_n)^{-1} \mathbf{A}^T \mathbf{b} \quad (2.37)$$

If \mathbf{A} is replaced with its SVD, and the equality $\mathbf{I}_n = \mathbf{V}\mathbf{V}^T$ is used, this can be rewritten as:

$$\hat{\mathbf{x}} = (\mathbf{V}\Sigma\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T + \lambda^2\mathbf{V}\mathbf{V}^T)^{-1} \mathbf{V}\Sigma\mathbf{U}^T\mathbf{b} \quad (2.38)$$

$$= (\mathbf{V}\Sigma^2\mathbf{V}^T + \lambda^2\mathbf{V}\mathbf{V}^T)^{-1} \mathbf{V}\Sigma\mathbf{U}^T\mathbf{b} \quad (2.39)$$

$$= (\mathbf{V}(\Sigma^2 + \lambda^2\mathbf{I}_n)\mathbf{V}^T)^{-1} \mathbf{V}\Sigma\mathbf{U}^T\mathbf{b} \quad (2.40)$$

$$= \mathbf{V}(\Sigma^2 + \lambda^2\mathbf{I}_n)^{-1} \mathbf{V}^T \mathbf{V}\Sigma\mathbf{U}^T\mathbf{b} \quad (2.41)$$

$$= \mathbf{V}(\Sigma^2 + \lambda^2\mathbf{I}_n)^{-1} \Sigma\mathbf{U}^T\mathbf{b} \quad (2.42)$$

$$\cdot \quad (2.43)$$

This can be rewritten as a summation using the left and right singular vectors to obtain:

$$\mathbf{A}^\dagger = \sum_{i=1}^{rank(\mathbf{A})} \mathbf{v}_i f_i \frac{1}{\sigma_i} \mathbf{u}_i^T \quad (2.44)$$

where the “filter factors” $f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$ have been introduced [78]. For values $\sigma_i \gg \lambda$, $f_i \simeq 1$. When $\sigma_i \ll \lambda$, however, $f_i \simeq \frac{\sigma_i^2}{\lambda^2}$. Thus, the introduction of the regularization matrix, and associated selection of the parameter λ have compensated for the ill-posed nature of the forward operator by effectively damping out the effect of those singular vectors with associated singular values below some level. For typical cases where exact noise levels are

unknown, methods such as the L-Curve of generalized cross validation (GCV) method are used for selection of an appropriate value for λ [71, 79].

2.3.2 L-Curve Analysis

For the problem of selecting a single regularization parameter, a number of methods have been developed over the years. Two common approaches are the L-Curve and the generalized cross validation, or GCV, method [71, 79]. The L-Curve method receives its name from the characteristic shape of the plots that are generated in the course of parameter selection [71]. The first step in using the L-Curve is to obtain inverse solutions to the problem using a range of regularization parameters. Typically, this range will span many orders of magnitude, and often the values of λ are logarithmically spaced. For each reconstruction, the residual norm is computed, as well as the norm of the solution times the regularization matrix. When these points are then plotted on a log-log graph, the resulting plot will frequently be shaped approximately like an L, as seen in Figure 2.1. The appropriate regularization parameter is then chosen to be the value whose point on the plot lies closest to the corner of the L-Curve. This decision is justified as representing a point at which minimization of the data fitting and regularization terms are best balanced. Automatic selection of this point can be done using methods based on locating the point of maximum curvature, or alternately finding the point which minimizes the distance to some “origin” [71].

L-Curve Analysis for Constrained Problems

The original theory for L-Curve analysis was performed in the context of unconstrained minimization. Fluorescence imaging, however, is a constrained problem because negative fluorescence values are non-physical. When solving such problems using unconstrained approaches such as the conjugate gradient (CG) method or LSQR, it is common to simply truncate the final solutions to set all negatively valued voxels to zero. When

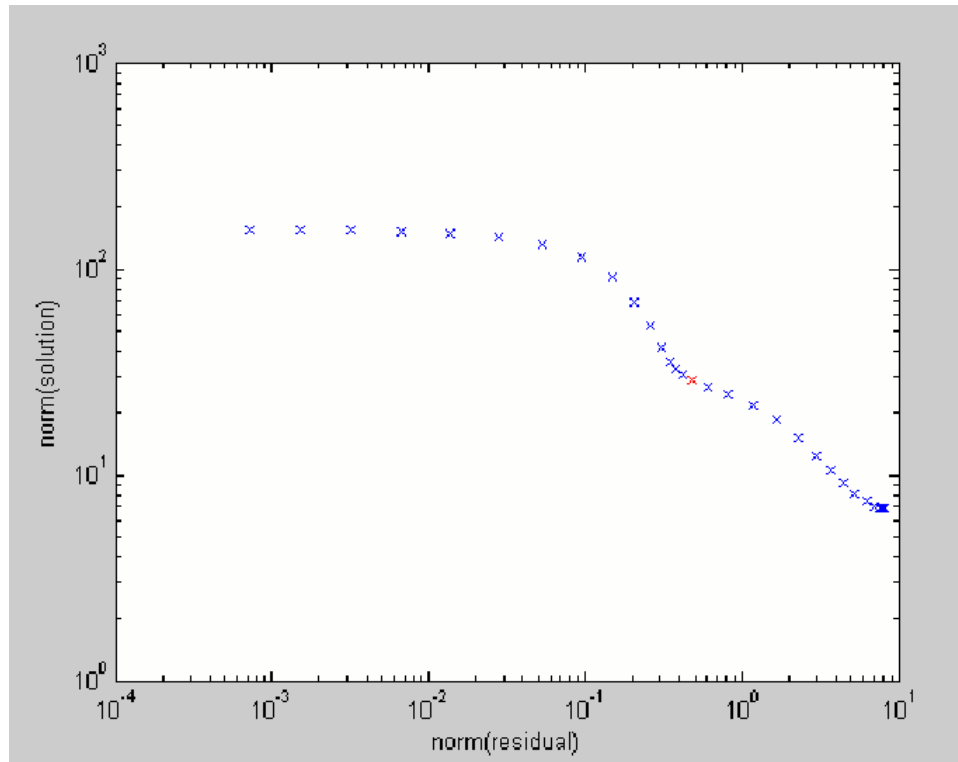


Figure 2.1: **Example L-Curve** Example of an L-Curve, computed from a non-contact FMT problem. This curve displays the characteristic L shape, although it is not as clearly defined as in some inverse problems. Each point plots the residual norm (X-axis) and solution norm (Y-axis) for a particular regularization parameter λ . As λ increases, so does the norm of the residual, while the norm of the solution decreases. Thus the plot is traversed from lower right to upper left as λ increases. The optimal value of lambda is chosen to be at the corner of the curve, as indicated by the red point in the plot.

these truncated solutions are then used to construct an L-Curve, an interesting behavior emerges.

Rather than the standard L-Curve structure, a log-log plot of the norm of the residual against the norm $\|\mathbf{L}\mathbf{x}\|$ results in the red line seen in Fig 2.2. Plotted alongside this is in blue a standard L-Curve, computed using the untruncated solutions. On the left half of the plot, corresponding to an increased value of λ , the two lines coincide with one another. As the plots pass the corner of the L-Curve, however, they diverge. As λ decreases, the standard L-Curve continues to see reductions in the overall residual, at the cost of increasing $\|\mathbf{L}\mathbf{x}\|$. Using the truncated solutions, however, the norm of the residual rapidly increases, leaving a curve whose minimum lies approximately at the corner of the standard L-Curve.

Similar to the normal L-Curve, this behavior can likely, given further study and analysis, be explained using the singular value decomposition, although we have not completed formal proofs of this. As the singular values σ_i decrease, the number of zero crossings of the associated singular vectors increases. That is, the singular vectors become more oscillatory. In a perfectly consistent system, with zero noise, the correlation between these singular vectors and the data vector being inverted will be very low. Noise introduced into the system, either in the form of model mismatch or measurement error, will have a higher correlation with these oscillatory singular vectors than the unperturbed solution. Because these vectors are highly oscillatory, incorporating them into the solution to any significant degree will tend to push portions of the total solution below zero. These components of the resulting solution are, however, critical to the further reduction of the residual norm. When the solutions are truncated, the negative values responsible for the reduction of the residual are eliminated, thereby allowing the overall residual to increase.

While the solutions that result from this approach are likely sub-optimal as compared to solutions obtained with explicitly constrained methods, this result does suggest

as useful combination of the two. Because of the difficulty implementing non-negativity constraints when using Krylov subspace approaches such as conjugate gradient and LSQR, many iterative constrained approaches use modifications of the steepest descent method. Unfortunately, because searches are often made along the same direction repeatedly as the algorithm descends the cost function, the number of iterations required to reach convergence can be much higher than with Krylov subspace methods. Using this modified L-Curve approach, however, it would be possible to use unconstrained methods as a method to quickly obtain a solution to initialize an explicitly constrained solution. By starting the algorithm closer to the true constrained minimum, the required number of steepest descent iterations could be greatly reduced, allowing for an overall reduction in computation time. At the moment, however, this remains largely speculation, and will require further analysis to prove such performance gains in practice.

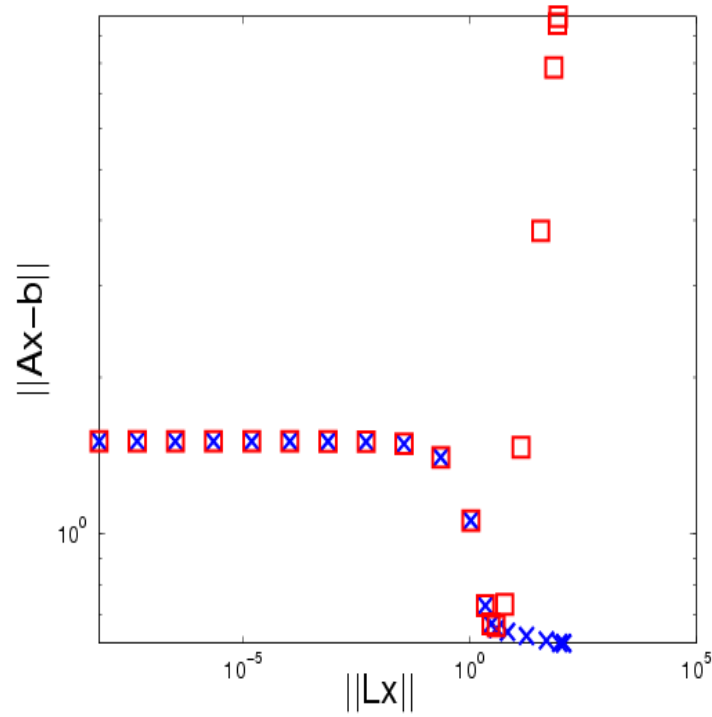


Figure 2.2: **L-Curve for a non-negatively constrained problem:** Computing an L-Curve prior to truncation of solutions results in the plot denoted with blue X's. If the same L-Curve approach is used on the solutions after truncation, the plot denoted with red squares results. Note that the inflection point of the red curve is as the same point as the corner of the standard L-Curve. This allows for automated selection of regularization parameters when using non-constrained methods to solve constrained problems.

Chapter 3

Statistical Modeling of the Born Ratio

3.1 Introduction

While the Born ratio provides an extremely useful tool for examining fluorescence signals, it simultaneously introduces a number of difficulties. In each of the raw received signals, the signal intensity is directly proportional to the signal noise, and higher signals have a proportionally larger influence upon the reconstructions. However, when the ratio of two such signals is taken, the resulting value does not behave like the intensities of its constituents. Thus, a higher ratio value could result from dividing two very high signals, but could just as easily be the division of two very low, and thus very noisy, signals. Moreover, because of the lower signal to noise ratios present in low intensity signals, the Born ratio values obtained can be highly unstable, resulting in values which can corrupt the final reconstruction.

To compensate for this shortcoming, previous work using the Born ratio relied on the use of thresholding to eliminate those data values whose signal to noise ratio was low enough that they could potentially lead to corrupted Born ratio values. This process was, however, highly subjective, and required manual tuning on the part of the investigator performing the inversion. This led to a high degree of variability in the reconstruction obtained from the same data set by different individuals. We will now detail an alternative method, which examines the statistical properties of the Born ratio, and seeks to use all collected data to form an approximate maximum likelihood solution. The resulting

algorithm uses a fixed point iteration and the conjugate gradient algorithm to obtain a solution. A second method, simplified based on empirical observations, eliminates an additional term from the cost function, resulting in a diagonally weighted least squares problem which can be solved using any standard algorithm for least squares problems.

3.1.1 Signal Modeling

Our initial goal was to determine an appropriate statistical model for the signals output from the CCD camera. A CCD such as that used in the FMT camera system makes measurements of the number of photons emitted from the surface of the target. Classically, photon counting processes have been modeled as Poisson random processes [80, 81]. For large numbers of photons (i.e. a large collection time with respect to the rate parameter), the Poisson process can be modeled as a Gaussian random variable with variance equal to its mean [82]. To determine the appropriateness of the Gaussian approximation, a simple experiment was performed.

A laser (672nm, 30mW) was used to illuminate one side of a 1.3cm thick chamber filled with a solution of Intralipid and ink with optical parameters $\mu_a = 0.58$ and $\mu_s = 10$. This resulted in a circular spot appearing on the collection side of the chamber. A series of 100 images were then collected, to increase the number of sample points used for statistical estimation. The location of the source with respect to the CCD was determined as the center of mass of all data points above some minimal threshold, selected to yield a large circular region on the CCD.

With the center determined, a number of bins were defined, based on radial distance from the center. Given that the single source was illuminating a homogeneous medium, diffusion theory states that the received signal on the opposite side of the slab should be dependent only upon the radial distance from the source location. Thus, by creating these bins, data points were effectively grouped as samples of theoretically identical random variables. Examining histograms of the data points collected into each bin revealed that

they did appear to be samples of a Gaussian random variable. As such, means and standard deviations were computed, to determine the relationship between the two.

Theoretically, it should be possible to model the photon stream incident on the CCD as a Poisson process, so that for a sufficiently high rate parameter it can be approximated as a Gaussian random variable with variance equal to its mean. From a statistical point of view, however, a significant amount of non-ideal processes take place is done between the incident photon stream and the final value output of the CCD's A/D converter. The CCD has some realistic efficiency less than one, and thus counts only a large fraction of the photons. This count is then stored electronically within the cells of the CCD, which are read out, row by row, through an analog amplifier. This amplifier scales the signal, which adjusts the dynamic range of the camera given the fixed input voltage range of the A/D converter. This scaling allows for low-light applications to make more use of the full 16-bit output range. Finally, the amplified voltage is passed to an A/D converter, which digitizes the signal. To achieve the differential increase in voltage needed to increment the digital output of the A/D requires that multiple additional photons be collected in the CCD cell. Thus the digitization itself has a scaling effect upon the signal. Additionally, in the course of the analog processing, a non-zero amount of additive, white, Gaussian read noise is added. Ultimately, given an input Poisson process approximated as a Gaussian $N(\mu, \mu)$, we derive a model $N(m, \sigma^2(m))$ for the signal at the output of the A/D with:

$$m = \frac{\varphi\mu\zeta}{\alpha} \quad (3.1)$$

$$\sigma^2(m) = \frac{\varphi^2\mu\zeta^2}{\alpha^2} + \sigma_r^2 \quad (3.2)$$

where φ is the quantum efficiency of the CCD, ζ is the gain of the amplifier, α is the constant relating voltage to output increments on the analog to digital converter, and σ_r^2 is the read noise associated with the CCD. Attempting to model the received data using this model resulted in the green line drawn in Figure 3.1, which does not fit the collected data well at higher signal intensities. The data does however seem to obey a similar model, with the variance is equal to the read noise, plus a component which is

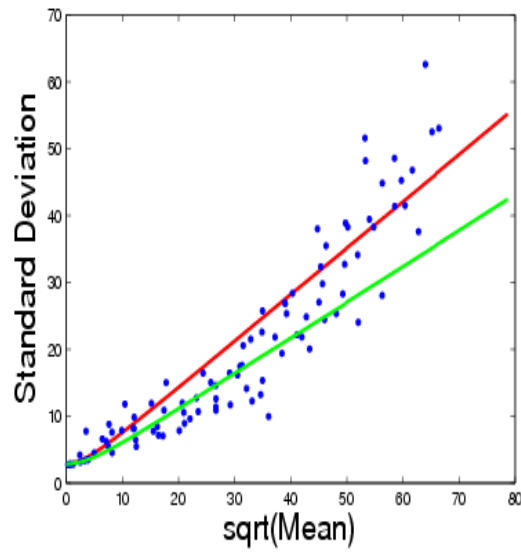


Figure 3.1: **Fit of Noise Model to Experimental Data** Scatter plot of experimentally collected noise data, along with a regression line from the model in (3.3). The R^2 value for the red fit line is 0.86. The green line shows the model as predicted based solely on published camera parameters. In order to better display the results at all intensity levels, the plot shows standard deviation plotted against the square root of the mean.

proportional to the mean signal. Thus, using the model:

$$\sigma^2(m) = \beta m + \sigma_r^2, \quad (3.3)$$

We obtained the parameters σ_r^2 and β directly from the data. This yielded a value of 2.8 for σ_r^2 , which is very close to the manufacturers stated value for the camera read noise. Applying a linear regression to the data, and taking into account the non-zero read noise, the line drawn in red in Figure 3.1 is obtained, yielding a value of $\beta = 0.4915$. This is approximately $1.5\times$ the modeled value of $\beta = 0.31$, obtained from camera specifications and use of (3.1), and suggests that either the values used to compute the modeled value may be incorrect, or there may be an additional source of noise in the system. In either case, for the remainder of this work we use the value of $\beta = 0.4915$, as it fits the collected data.

In order to write an expression for the joint density of the fluorescence and excitation measurements, so that we can then derive an expression for the density of their ratio, their covariance needs to be modeled both across space at a fixed wavelength, and also across wavelengths. In the work described here, we make the simplifying assumptions that the measurements are uncorrelated both in space and across wavelengths. This results in a joint density for the received signals of:

$$\begin{bmatrix} \mathbf{u}_{fl} \\ \mathbf{u}_{exc} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{m}_{fl} \\ \mathbf{m}_{exc} \end{bmatrix}, \text{diag} \left(\beta \begin{bmatrix} \mathbf{m}_{fl} \\ \mathbf{m}_{exc} \end{bmatrix} + \sigma_r^2 \right) \right) \quad (3.4)$$

While these independence assumptions are not physically accurate, they greatly simplify the process of finding an inverse solution. As we will show in the results section, this assumption does not prevent useful results from being obtained in practice.

3.1.2 Density Derivation

To derive a density for the Born ratio from the experimentally determined signal densities, we begin with the definition of the Born ratio:

$$U_{born} = \frac{U_{fl}}{U_{exc}} \quad (3.5)$$

where U_{born} is the Born ratio data, and U_{fl} and U_{exc} are a corresponding pair of fluorescence and excitation measurements, respectively. While both of the individual signals are Gaussian, the generalized ratio of two Gaussians is not. In the case of zero mean, unit variance Gaussians, the resulting ratio will have a Cauchy density. For the more general case, a search of the literature provided the density for two correlated normal random variables [83]:

$$p(\mathbf{u}_i) = \frac{b(\mathbf{u}_i)d(\mathbf{u}_i)}{\sqrt{2\pi}\sigma(m_{fl,i})\sigma(m_{exc,i})a(\mathbf{u}_i)^3} \left[\Phi \left\{ \frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \right\} - \Phi \left\{ -\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \right\} \right] + \frac{1}{\pi\sigma(m_{fl,i})\sigma(m_{exc,i})a(\mathbf{u}_i)^2} e^{-0.5c} \quad (3.6)$$

where \mathbf{u}_i are the individual Born ratio measurements, and $a(\mathbf{u}_i)$, $b(\mathbf{u}_i)$, c and $d(\mathbf{u}_i)$ are defined as:

$$\begin{aligned} a(\mathbf{u}_i) &= \left(\frac{\mathbf{u}_i^2}{\sigma^2(m_{fl,i})} + \frac{2\rho\mathbf{u}_i}{\sigma(m_{fl,i})\sigma(m_{exc,i})} + \frac{1}{\sigma^2(m_{exc,i})} \right)^{1/2} \\ b(\mathbf{u}_i) &= \frac{m_{fl,i}\mathbf{u}_i}{\sigma^2(m_{fl,i})} - \frac{\rho(m_{fl,i} + m_{exc,i}\mathbf{u}_i)}{\sigma(m_{fl,i})\sigma(m_{exc,i})} + \frac{m_{exc,i}}{\sigma^2(m_{exc,i})} \\ c &= \frac{m_{fl,i}^2}{\sigma^2(m_{fl,i})} - \frac{2\rho\sigma(m_{fl,i})\sigma(m_{exc,i})}{\sigma(m_{fl,i})\sigma(m_{exc,i})} + \frac{m_{exc,i}^2}{\sigma^2(m_{exc,i})} \\ d(\mathbf{u}_i) &= \exp \left\{ \frac{b^2(\mathbf{u}_i) - ca^2(\mathbf{u}_i)}{2(1 - \rho^2)a^2(\mathbf{u}_i)} \right\}. \end{aligned} \quad (3.7)$$

with ρ being the correlation coefficient relating the two variables and $\Phi(u)$ the usual integral of the unit normal density $\phi(\mathbf{u}_i)$:

$$\begin{aligned} \Phi(u) &= \int_{-\infty}^u \phi(y)dy \\ \phi(u) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \end{aligned} \quad (3.8)$$

The independence assumption made in constructing our signal model means that the correlation coefficients will be zero. Given this assumption and some algebraic rearrangement, the various functional components can be rewritten as:

$$\begin{aligned}
 a(\mathbf{u}_i) &= \left(\frac{\mathbf{u}_i^2}{\sigma^2(m_{fl,i})} + \frac{1}{\sigma^2(m_{exc,i})} \right)^{1/2} \\
 b(\mathbf{u}_i) &= \frac{m_{fl,i}\mathbf{u}_i}{\sigma^2(m_{fl,i})} + \frac{m_{exc,i}}{\sigma^2(m_{exc,i})} \\
 c &= \frac{m_{fl,i}^2}{\sigma^2(m_{fl,i})} + \frac{m_{exc,i}^2}{\sigma^2(m_{exc,i})} \\
 d(\mathbf{u}_i) &= \exp \left\{ -0.5 \frac{m_{exc,i}^2 \left(\frac{m_{fl,i}}{m_{exc,i}} - \mathbf{u}_i \right)^2}{(\sigma^2(m_{exc,i})\mathbf{u}_i + \sigma^2(m_{fl,i}))} \right\}.
 \end{aligned} \tag{3.9}$$

This density provides a basis for further derivation. To obtain an inverse solution, we look to use a maximum likelihood (ML) approach, which uses the log-likelihood function to convert a product of probabilities into a sum of probabilities. The current formulation of the density, with two separate terms, makes this significantly more difficult, so we seek now to determine a simplified version of the density which consists of a single term.

The first simplification made is to eliminate terms containing Φ by making the approximation:

$$\left[\Phi \left\{ \frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \right\} - \Phi \left\{ -\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \right\} \right] \simeq 1. \tag{3.10}$$

As is shown in Appendix A.1, this approximation holds with probability $(1 - 2 \times 10^{-6})$ when either $m_{fl,i} > 18$ or $m_{exc,i} > 18$. Note that these required signal levels are less than 0.3% of the maximum output values of $2^{16} = 65536$. Additionally, the introduction of these thresholds does not run counter to the goal of eliminating user defined thresholds because these are static and analytically determined.

Furthermore, in Appendix A.2, we show that by slightly increasing the above thresholds to $m_{fl,i} > 18$ and $m_{exc,i} > 20$, the second term in the expression will be smaller than the first by several orders of magnitude. Thus, we simplify the overall density by eliminating the low magnitude term, resulting in an approximate density for the

Born ratio data of:

$$p(\mathbf{u}_i) \simeq \left[\frac{b(\mathbf{u}_i)d(\mathbf{u}_i)}{\sqrt{2\pi}\sigma(m_{fl,i})\sigma(m_{exc,i})a(\mathbf{u}_i)^3} \right] \quad (3.11)$$

This density, simplified to a single term, is now in a form from which an approximate maximum likelihood solution can be constructed.

3.1.3 Density Conditioning

At this point, the density function given above is a function of the collected data, $m_{fl,i}$ and $m_{exc,i}$, the mean values for each of the individual excitation and fluorescence signals. In order to construct an inverse problem, we must first relate this density to the forward model \mathbf{W} and \mathbf{x} , the image we are attempting to recover. The parameter $m_{exc,i}$ is the mean value of the excitation channel for the i^{th} measurement and is therefore independent of the fluorochrome distribution. Explicit computation of this value requires knowledge about the optical absorption and scattering parameters of the medium, and a model which uses these to accurately predict measurement values. This is the problem of traditional diffuse optical tomography (DOT). Avoiding direct solution of this problem is one of the advantages of using the Born ratio, and as such, we do not want to introduce it here simply to determine parameter values. Instead, we use the approximation:

$$m_{exc,i} \simeq u_{exc,i}. \quad (3.12)$$

Thus we simply use the received signals as an approximation of the mean value. Next, we assume that the forward model for the Born ratio data generates the ratio of the mean signal values to write:

$$\mathbf{W}_i \mathbf{x} = \frac{m_{fl,i}}{m_{exc,i}}. \quad (3.13)$$

Here, we are assuming that the values generated by the forward model are equivalent to the ratio of the means of the two distributions. This equation can then be rearranged to write an expression for $m_{fl,i}$ as:

$$m_{fl,i}(\mathbf{x}) \simeq u_{exc,i} \mathbf{W}_i \mathbf{x}. \quad (3.14)$$

Equations (3.12) and (3.14) now allow the density for the Born ratio data to be rewritten in a form conditioned on the image:

$$p(\mathbf{u}_i|\mathbf{x}) = \left[\frac{b(\mathbf{u}_i, \mathbf{x})d(\mathbf{u}_i, \mathbf{x})}{\sqrt{2\pi}\sigma(m_{fl,i}(\mathbf{x}))\sigma(m_{exc,i})a(\mathbf{u}_i, \mathbf{x})^3} \right], \quad (3.15)$$

which enables the development of a statistical inverse solution to proceed.

3.1.4 Maximum Likelihood Solution

An inverse solution to the fluorescence molecular tomography (FMT) problem can now be constructed as a maximum likelihood (ML) problem. The general form for an ML problem is:

$$\hat{\mathbf{x}}_{ML} = \arg \max_{\mathbf{x}} p(\mathbf{u}|\mathbf{x}). \quad (3.16)$$

Given the independence assumptions made in the previous section, the right side of (3.16) can be rewritten using Eq (3.15) as:

$$\hat{\mathbf{x}}_{ML} = \arg \max_{\mathbf{x}} \prod_i \left[\frac{b(\mathbf{u}_i, \mathbf{x})d(\mathbf{u}_i, \mathbf{x})}{\sqrt{2\pi}\sigma(m_{fl,i}(\mathbf{x}))\sigma(m_{exc,i})a(\mathbf{u}_i, \mathbf{x})^3} \right] \quad (3.17)$$

Taking the log-likelihood of this probability converts the product to a summation. Eliminating terms which are not dependent upon \mathbf{x} (they will be constant for purposes of the optimization) yields:

$$\hat{\mathbf{x}}_{ML} = \arg \max_{\mathbf{x}} \sum_i [\log(b(\mathbf{u}_i, \mathbf{x})) - \log(\sigma(m_{fl,i}(\mathbf{x}))) - 3\log(a(\mathbf{u}_i, \mathbf{x})) + \log(d(\mathbf{u}_i, \mathbf{x}))] \quad (3.18)$$

To simplify notation, we define a function:

$$g_i(\mathbf{u}_i, \mathbf{x}) = \log(b(\mathbf{u}_i, \mathbf{x})) - \log(\sigma(m_{fl,i}(\mathbf{x}))) - 3\log(a(\mathbf{u}_i, \mathbf{x})) \quad (3.19)$$

and a vector $\mathbf{q}(\mathbf{u}, \mathbf{x})$ with elements:

$$q_i(\mathbf{u}_i, \mathbf{x}) = \frac{m_{exc,i}}{\sqrt{\sigma(m_{exc,i})^2 \mathbf{u}_i + \sigma(m_{fl,i}(\mathbf{x}))^2}} \quad (3.20)$$

and obtain:

$$\hat{\mathbf{x}}_{ML} = \arg \max_{\mathbf{x}} \sum_i \left[g_i(\mathbf{u}_i, \mathbf{x}) - 0.5 \left(q_i(\mathbf{u}_i, \mathbf{x}) \left(\frac{m_{fl,i}(\mathbf{x})}{m_{exc,i}} - \mathbf{u}_i \right) \right)^2 \right]. \quad (3.21)$$

The second term of the above equation can be rewritten into matrix form, yielding:

$$\hat{\mathbf{x}}_{ML} = \arg \max_{\mathbf{x}} \left[\sum_i g_i(\mathbf{u}_i, \mathbf{x}) - 0.5 \|\mathbf{Q}(\mathbf{u}, \mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2 \right] \quad (3.22)$$

where the matrix \mathbf{Q} is defined as: $\mathbf{Q}(\mathbf{u}, \mathbf{x}) = \text{diag}(\mathbf{q}(\mathbf{u}, \mathbf{x}))$.

To solve this maximization equation we want to find the point at which the gradient of the right side is equal to zero, as this corresponds to a maximum in the function. Defining the cost function to be maximized as $f(\mathbf{x})$, the results of Appendices A.3.2 and A.3.1 can be used to write:

$$\frac{df(\mathbf{x})}{d\mathbf{x}} = \left[\sum_i \frac{dg_i(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} \right] - \mathbf{W}^T \mathbf{Q}(\mathbf{x})^T \mathbf{Q}(\mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u}) - \left(\frac{d\mathbf{q}(\mathbf{x})}{d\mathbf{x}} \right)^T A(\mathbf{x})^T A(\mathbf{x}) \mathbf{q}(\mathbf{u}, \mathbf{x}) \quad (3.23)$$

where we have defined $A(\mathbf{x})$ in App. A.3.1 to be:

$$A(\mathbf{u}, \mathbf{x}) = \text{diag}(\mathbf{W}\mathbf{x} - \mathbf{u}). \quad (3.24)$$

Setting equation (3.23) to zero and rearranging yields:

$$\mathbf{x} = [\mathbf{W}^T \mathbf{Q}(\mathbf{x})^T \mathbf{Q}(\mathbf{x}) \mathbf{W}]^{-1} \times \left[\sum_i \frac{dg_i(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} + \mathbf{W}^T \mathbf{Q}(\mathbf{x})^T \mathbf{Q}(\mathbf{x}) \mathbf{u} - \left(\frac{d\mathbf{q}(\mathbf{x})}{d\mathbf{x}} \right)^T A(\mathbf{x})^T A(\mathbf{x}) \mathbf{q}(\mathbf{x}) \right]. \quad (3.25)$$

Given the dependence on the solution \mathbf{x} of the terms on the right, this result suggests a fixed point iteration, which, calling the right hand side $\zeta(\mathbf{x})$, can be written as:

$$\mathbf{x}^{(n+1)} = \zeta(\mathbf{x}^{(n)}). \quad (3.26)$$

To solve this fixed point iteration, we use the conjugate gradient algorithm, given that the matrix being inverted is symmetric and positive definite. To initialize the algorithm and obtain a value for x_0 , we set a fixed threshold, based entirely upon the

excitation signal, such that only 25% of the data is retained. We then solve the thresholded inversion problem using either ART or LSQR, and use the resulting solution as the value of x_0 .

Reexamining equation (3.22), it is clear that the cost function could be simplified significantly by the elimination of first term within the brackets. Given empirical evidence, this term is in general several orders of magnitude smaller than the second term, and consequently, can be removed from the optimization with minimal effect. This leaves:

$$\begin{aligned}\hat{\mathbf{x}}_{ML} &= \arg \max_{\mathbf{x}} [-0.5 \|\mathbf{Q}(\mathbf{u}, \mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2] \\ &= \arg \min_{\mathbf{x}} \|Q(\mathbf{u}, \mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2\end{aligned}\tag{3.27}$$

which is just a weighted least squares problem. However, the weighting matrix \mathbf{Q} is still dependent upon the value of \mathbf{x} . To eliminate this, we apply a point estimate similar to that used earlier to obtain μ_{exc} . Replacing $m_{fl,i}$ with $u_{fl,i}$ leaves:

$$\hat{\mathbf{x}}_{ML} = \arg \min_{\mathbf{x}} \|Q(\mathbf{u})(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2,\tag{3.28}$$

which can be solved with any standard least squares solver.

3.1.5 Results and Discussion

Testing of this statistical solution was done using the planar FMT system described in Section 1.6.1. A grid of 13×13 virtual detector locations was defined, corresponding to uniform spacing over a $2.2 \times 2.2 \text{ cm}^2$ region centered around the source array, on the opposite side of the illuminated volume from the sources. Individual detector values were obtained by integrating the values of CCD pixels corresponding to 1mm diameter equivalent circular detectors on the camera. The solution space, defined to be the $2.4 \times 2.4 \times 1.3 \text{ cm}^3$ region centered around the sources, was discretized into $20 \times 20 \times 21 = 8400$ volume elements (voxels), each of dimension $0.12 \text{ cm} \times 0.12 \text{ cm} \times 0.0619 \text{ cm}$. On all reconstruction images, this region is demarcated by a red box. For all cases, the imaging

chamber was filled with a solution of 1% intralipid and 0.5% India ink which has optical properties $\mu_a = 0.58\text{cm}^{-1}$ and $\mu_s = 10\text{cm}^{-1}$. These values were chosen to correspond closely with the bulk background optical parameters of mice.

Phantom Experiment The first set of data employed came from phantom measurements, providing a simple scenario to validate this method. The phantom consisted of two tubes, 8mm long and 2.5mm wide, each containing 400nM Cy5.5, suspended in a tank of intralipid solution. The two tubes were placed 5mm apart along the lateral axis, with one tube located against the detector side of the volume, and the other in its center, corresponding to distances of 0.6cm and 1.2cm from the source plane. Figures 3.2(a) and 3.2(d) show flat images (two dimensional photographs) at the fluorescence and excitation wavelengths, respectively. They were taken prior to the chamber being filled with intralipid, and while the objects appear to be at the same distance from the camera, they are in fact separated by 0.6cm along the source-camera axis.

Euthanized Mouse The second set of data examined consisted of two plastic catheter tubes with an inner diameter of 0.8mm, containing 1 μ M Cy5.5, implanted in the body of an euthanized mouse. As can be seen in the fluorescence reflectance and flat images (Figures 3.3(a) and 3.3(d) respectively), one tube was inserted down the esophagus while the other was implanted subcutaneously on the anterior side of the torso. The goal of this experiment was to provide an easily identifiable and quantifiable target (the tubes) under conditions where the background optical parameters are similar to those of an *in vivo* experiment.

In-vivo Mouse The third study presented considered an *in vivo* experiment. A Her2-neu transgenic mouse was selected, exhibiting a spontaneous breast tumor. The tumor was located in the left mammary pad as indicated by an arrow on Fig 3.4(a), and was approximately 5x6mm in size, as measured externally with calipers. The an-

imal was injected with 2nmol of the ProSense680 cathepsin-sensitive activatable probe (VisEn Medical, Woburn MA) twenty four hours prior to imaging. Accompanying the reconstructions is a flat image at the excitation wavelength in 3.4(a), on which the tumor is visible as a dark region in the left mammary pad (right side of image).

Description of Results For each data set, we present and compare two methods. The first method obtains a solution using the normalized Born ratio after thresholding to reject all ratio data with an excitation wavelength intensity less than a fixed value. The threshold values used for each data set were determined by obtaining inverse solutions for thresholds set between the 0th and 95th percentile of the excitation wavelength measurements, in 5% increments. The ‘optimal’ solution was then chosen from among these 20 results, based on prior knowledge of the target being imaged. Results with the chosen thresholds varied from 15% to 35% of the Born ratio values being retained for use in the inversion algorithm.

The second method utilized the statistical methods derived in Section 3.1.4. The previously established threshold requiring $u_{exc,i} > 20$ counts or $u_{fl,i} > 18$ counts was applied to the data prior to inversion, resulting in between 0.1% and 32% of the data being removed. For both reconstruction methods, depth dependent regularization was achieved by making the stopping point of the iterations increase exponentially as a function of depth, from either 4 or 10 iterations, which optimally reconstructed surface activity, to 25 for the middle reconstructed slice. Further iterations resulted in only minor changes in the reconstruction. We note that compared to utilizing a depth independent number of iterations, the depth dependent regularization employed does not alter the relative performance of the statistical vs. the automatic method but yields a more accurate reconstruction of the physical dimensions of the fluorescence activity, as was also noted in [84].

Comparison Metrics In addition to subjective visual analysis of results, we

also used two analytic metrics to compare my statistical method to existing threshold based methods. These metrics serve to quantify the size and location of the resolved object, to compare to known values.

In order to compute these two metrics, we first needed to determine the extent of the reconstructed object. Given the differences in intensity owing to the different depths at which fluorescent objects are located, we started this process by manually selecting, for each object, a slice which contained a clear cross section of that object. We then took the voxel in that slice with the highest amplitude to be the initial location of the detected object. Selecting a threshold of 50% of this initial value, we allowed the object to grow by iteratively incorporating neighboring voxels in 3D whose amplitude were above the threshold. In this manner, we determined which voxels were considered to lie within the resolved object.

The two metrics used to quantify the reconstructions were the centroid of the reconstructed object, and its average dimensions. The centroid was computed as the mean of the objects' constituent voxel locations, weighted by the associated image intensities. Average dimensions were determined by first computing the average location of each side of the object, then taking the distance between the two sides associated with each of the X, Y, and Z axes.

3.1.6 Phantoms in Intralipid

In addition to the reconstructions for the phantom experiment, Figure 3.2 displays two images of the phantoms without intralipid present. Figure 3.2(d), the flat image, is a front illuminated image of the chamber, while Figure 3.2(a), the fluorescence reflectance, is an image captured at the fluorescence wavelength while front illuminating the chamber with light at the excitation wavelength. Note that the reconstruction slices displayed in Figure 3.2 are chosen at the depth of maximum intensity in the statistical reconstruction, and were obtained using 4 iterations for the surface object, and 25 for the central one.

Data Set	Object	Inversion Method	Centroid Location	Average Dimensions
Phantom	Central	Threshold	(-0.53,-0.18,0.55) cm	$0.47 \times 0.66 \times 0.71$
		Statistical	(-0.54,-0.20,0.53) cm	$0.45 \times 0.73 \times 0.69$
		Actual Value	(-0.51,-0.23,0.6) cm	$0.48 \times 0.8 \times 0.48$
	Detector Side	Threshold	(0.56,-0.18,1.15) cm	$0.27 \times 0.62 \times 0.26$
		Statistical	(0.55,-0.03,1.18) cm	$0.27 \times 0.56 \times 0.21$
		Actual Value	(0.57, -0.23, 1.2) cm	$0.48 \times 0.8 \times 0.48$
Euthanized	Esophageal	Threshold	(0.05,-0.17,0.64) cm	$0.28 \times 0.53 \times 0.59$
		Statistical	(0.06,-0.06,0.61) cm	$0.36 \times 1.35 \times 0.76$
		Actual Values	Unknown	$0.2 \times 1.5 \times 0.2$
	Subcutaneous	Threshold	(0.60,-0.21,1.14) cm	$0.34 \times 0.61 \times 0.24$
		Statistical	(0.66,-0.14,1.12) cm	$0.30 \times 0.56 \times 0.30$
		Actual Values	Unknown	$0.2 \times 1.0 \times 0.2$
<i>In Vivo</i>	Tumor	Threshold	(0.34,0.20,1.22) cm	$0.20 \times 0.63 \times 0.13$
		Statistical	(0.56,0.32,1.14) cm	$0.32 \times 0.59 \times 0.28$

Table 3.1: **Locations and Sizes of Reconstructed Objects** Centroid and Maximum Dimension are given in (x,y,z) coordinates except for the subcutaneous tube in the euthanized mouse, where width, length, depth of the tube are given, as it is not parallel with any of the three primary axes. Measurements on the reconstructions are computed as described in the text. All size measurements are given in cm.

Using the manually optimized method, the threshold was set to retain 15% of the data for use in the reconstruction. Figures 3.2(b) and 3.2(c) depict two coronal slices of the reconstructed three-dimensional image. In each slice, the solution clearly shows a fluorescing object present at the location of the phantoms. The slice located deeper within the volume, shown in Figure 3.2(b), resolves the object to be larger in size compared to the slice in 3.2(c), a result of decreasing resolution at greater depths.

The results with our statistical inversion are illustrated in Figures 3.2(e) and 3.2(f), where very similar reconstructions are seen. The object in Figure 3.2(e) is slightly larger than that in 3.2(b), but remains consistent with the size of the phantoms seen in the flat and fluorescence images. Figure 3.2(f) shows an image very similar to that in Figure 3.2(c). These images were achieved with the use of 99.9% of the collected data. Only about one hundred of the 14000 data points collected did not satisfy the minimum signal requirements.

The sizes reconstructed with the two methods are reasonably consistent with one another. Notably, in both cases, the resolution is poorest along the Z-axis between the source and detector planes. Both methods overestimate the object size along this dimension. Additionally, while the thickness (Z-axis measurement) of the central tube is overestimated, that of the detector side tube is underestimated. This variation in reconstructed object thickness is seen for both methods, with all data sets, and is due to the restrictions on source and detector locations when using a slab geometry.

3.1.7 Euthanized Mouse

As with the tubes in intralipid, flat and reflectance fluorescence images are displayed in Figures 3.3(d) and 3.3(a), alongside reconstructions corresponding to the slice at which the maximum intensity was obtained with the statistical reconstruction method. As with the tube phantom, these reconstructions use 4 iterations for the subcutaneous tube, and 25 for the esophageal one.

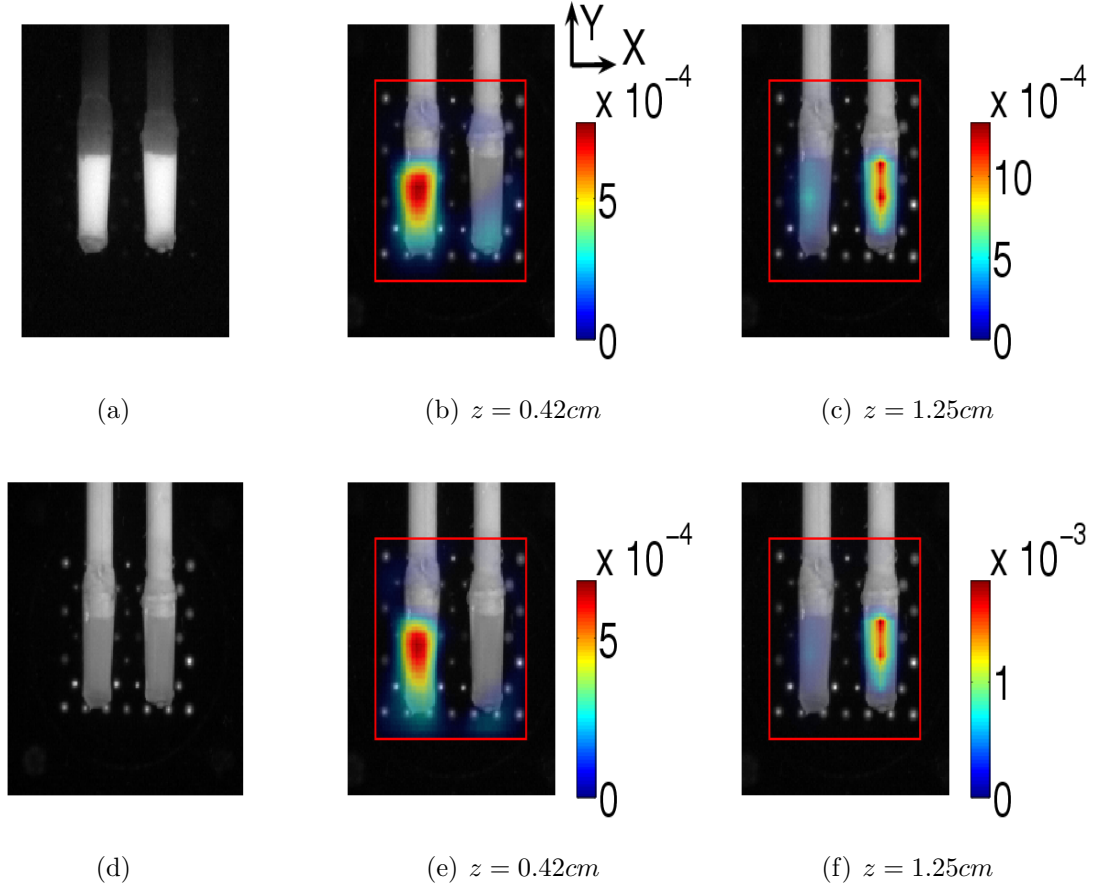


Figure 3.2: **Dual Tube Phantom:** (a) Reflectance Fluorescence image (b) Manual Threshold: Coronal slice through tube @ $z=0.42\text{cm}$ (c) Manual Threshold: Coronal slice through tube @ $z = 1.25\text{cm}$ (d) Flat (Photographic) Image (e) Fixed Point: Coronal Slice through $z=0.42\text{cm}$ (f) Fixed Point: Coronal Slice through $z=1.25\text{cm}$ In all images, $z = 0\text{cm}$ corresponds to the source side of the slab, with a total slab thickness of 1.3cm . The red box in the image denotes the boundary of the solution space, and the orientation of the coordinate system specified in 3.2(b) is shared by all images.

The results using the threshold method are shown in Figures 3.3(b) and 3.3(c). These were obtained by retaining 35% of the data. As with the phantom, two coronal slices are shown, one through each of the implanted tubes. Figure 3.3(c) shows a vaguely tube-like object located just under the surface. Its location corresponds well with the fluorescence image seen in Figure 3.3(a), but it is not nearly as large as would be expected. The object in the center of the volume can be seen in Figure 3.3(b). However, rather than appearing to be a uniform tube, the object appears to be more centrally concentrated.

Figures 3.3(b) and 3.3(c) contains the results for this data set using our fixed point iteration. The subcutaneous tube is resolved in a manner similar to that using the thresholding technique. It is, however, slightly larger, more tube-like, and more consistent with Figure 3.3(a). Looking at the esophageal tube, it is resolved to be a much more uniform tube than it was in Figure 3.3(b). For this data set, the minimum signal requirements resulted in 32% of the data being left out of the reconstruction, as with the threshold method a significantly higher number than for the phantom experiment. This seems to be a result of a higher level of absorption within the mouse body. While for the other two data sets, less than 200 excitation measurements failed the minimum criterion, for the euthanized mouse, nearly 3700 points failed the test. The additional absorption also seems to have affected the resulting fluorescence levels, as a significantly higher number of fluorescence measurements failed as well.

Observing the metrics in Table 3.1, we see that the esophageal tube, as with the central tube of the previous data set, is resolved to have a much larger X-Z cross section than the subcutaneous tube. However, the length (Y-dimension) of the esophageal tube is much more accurately resolved using the statistical reconstruction. While the statistical method does result in average X and Z dimensions that are slightly larger than the threshold method, the difference is small, and comparable to the size of a single voxel.

Looking at the subcutaneous tube, the location shift was less than a fraction of a voxel, while the sizes along the three dimensions are very similar between the two

methods. In both cases, while the width and depth are slightly overestimated, the length is underestimated, and the reconstruction does not appear to extend the full length of the tube seen in 3.3(a).

3.1.8 In-Vivo Mouse

Figure 3.4(a) shows a flat image of the mouse being imaging. In it, the tumor is clearly visible as a dark spot in the left mammary pad (right side of image). This is indicative of the increased angiogenesis and blood concentration expected in the tumor. This reconstruction used 10 iterations to reconstruct the tumor.

For this data set, the manually selected threshold was set to retain 25% of the data, based on results from previous studies [4]. The resulting inversion is seen in Figure 3.4(b). A fluorescing mass is clearly present on the anterior surface of the mouse, just to the right of center. Results with the fixed point iteration are shown in Figure 3.4(c). There, the tumor is resolved to be slightly larger than with the threshold, and its spatial extent is more consistent with the dark appearance of the tumor in the flat image. As with the phantom experiment, less than 0.1% of the data was dropped from the reconstruction.

Looking at the analytic metrics, we can see that the statistical method resolves the tumor to be slightly larger in width and depth than the threshold method. This is more consistent with the approximate size of the tumor of $5mm \times 6mm$, and represents a more ellipsoidal and less flattened tumor than the threshold method.

3.2 Conclusions

This work has presented a new method by which Fluorescence Molecular Tomography solutions can be obtained based on the normalized Born approximation. This method incorporates statistical properties of the measurements at the excitation and emission wavelengths in order to eliminate the use of operator defined thresholds. The method performed consistently and automatically across a variety of experimental scenar-

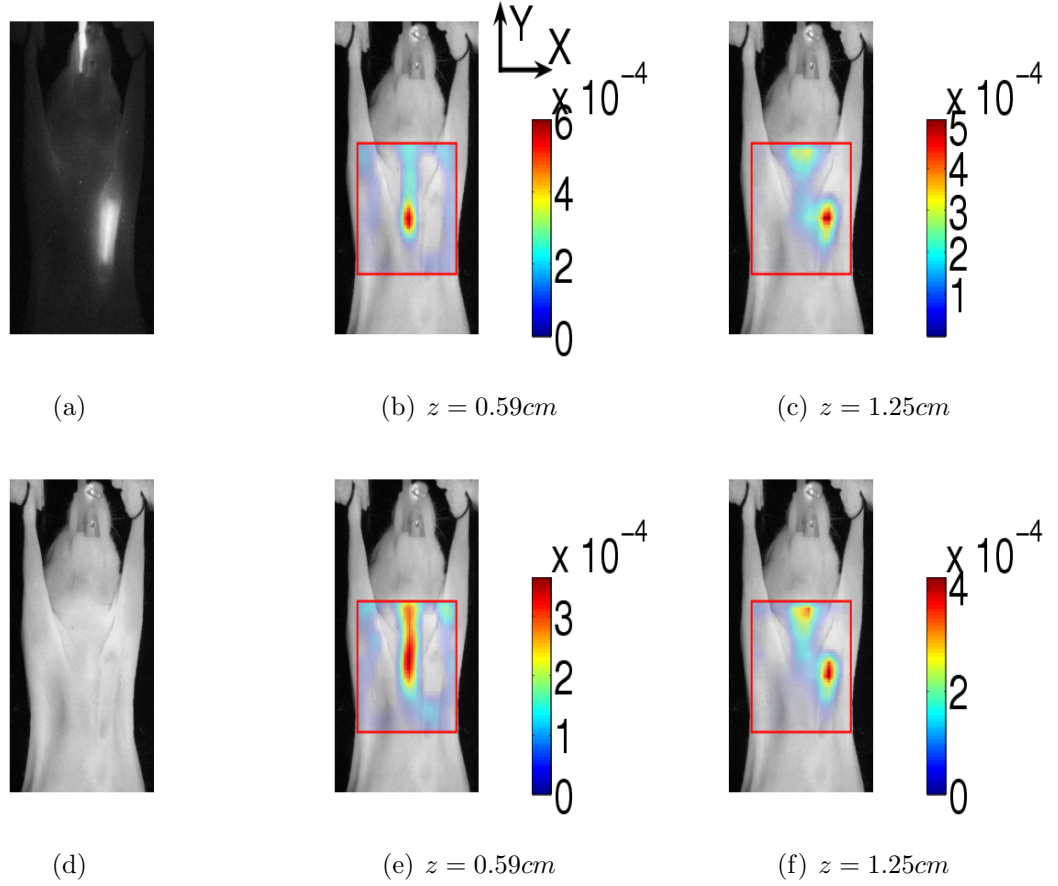


Figure 3.3: **Euthanized Mouse:** (a) Fluorescence Reflectance (b) Manual Threshold: Coronal slice through esophageal tube (c) Manual Threshold: Coronal slice through subcutaneous tube. (d) Flat (Photographic) Image (e) Fixed Point: Coronal slice through esophageal tube (f) Fixed Point: Coronal slice through subcutaneous tube. In all images, $z = 0cm$ corresponds to the source side of the slab, with a total slab thickness of $1.3cm$. The red box denotes the boundary of the solution space.

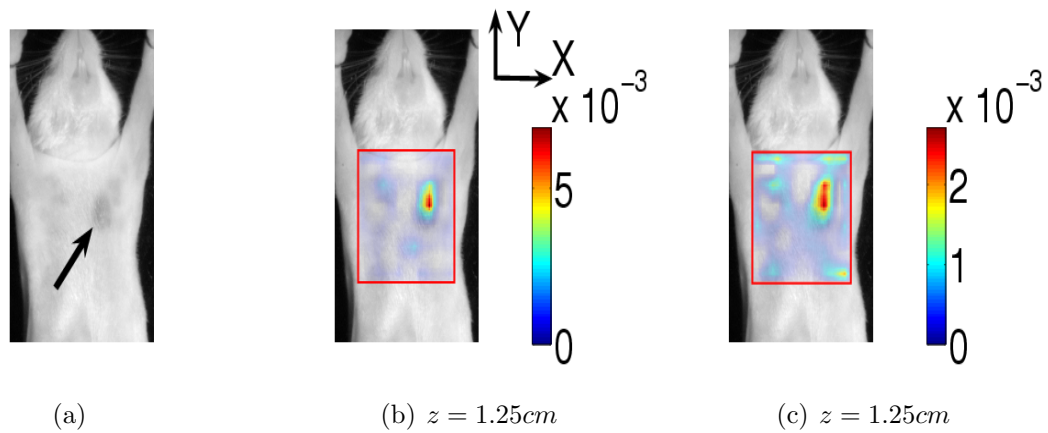


Figure 3.4: **In-Vivo experiment.** (a) Flat (Photographic) Image. Arrow indicates location of implanted tumor. (b) Manually Selected Threshold (c) Fixed Point Iteration. In both reconstructions, the red box denotes the boundary of the solution space.

ios, something which was not possible with previous thresholding techniques. Examining the results we see that they compare favorably with those obtained using empirically defined thresholds, even when the thresholds were tuned using knowledge of the correct result.

Possible future extensions to this work include the use of fully nonlinear inversion methods, more complex regularization methods, the inclusion of correlation between the various data points, and spatially variant regularization to improve the dependence of size estimates on depth.

One of the advantages of using repeated solution of linear systems, such as we have done here, is that such systems are in general efficient and easily implemented. If instead, we were to apply the conjugate gradient method directly to our nonlinear cost function in (3.22), we could eliminate the need for the fixed point iteration, at the cost of greater algorithmic and computational complexity.

Another possible extension of this work is the inclusion of correlation into the equations. Two independence assumptions were made to arrive at the weights detailed in this paper. The first of these was that each fluorescence measurement was independent of its associated excitation measurement. The second was that each fluorescence measurement was independent of every other fluorescence measurement, and likewise for the excitation measurements. Neither of these conditions is truly accurate. The intensity of excitation light reaching a detector is highly predictive of the intensity of fluorescence that detector may see, although the exact value seen is still primarily dependent upon the physical distribution of the fluorochrome.

Overall, the method presented offers an inversion scheme that can enable more robust tomographic performance, by reducing image accuracy dependencies on calibration studies and user defined thresholds. Guided by a-priori information based on the statistical nature of the measurements, this approach potentially offers a generalized and transferable inversion method across different platforms appropriate for standardization

of optical tomography techniques. Further studies will involve application of the method to a larger number of in-vivo data to confirm the validity of application across a large number of different potential targets, organs and tissues sizes.

Chapter 4

Effects of Parameter Selection on Finite Element Modeling for Normalized Fluorescence Imaging

4.1 Introduction

While non-contact, rotational FMT imaging systems such as described in §1.6.2 offer improvements in image quality over previous slab based systems, the ill-posed nature of the inverse problem still limits the resulting image resolution. Multi-modality approaches have been suggested as a method to help overcome this limitation [26,30,31,85]. Available from imaging modalities such as X-ray CT or MRI, structural prior information can be used in both modeling and inversion, and has been shown to significantly improve image quality. Multiple approaches have been proposed to incorporate prior information into the inverse problem, with methodologies that avoid the use of hard priors offering promising characteristics in order to avoid image bias [86].

Here we investigate the forward modeling aspect of the hybrid tomographic problem. Our imaging system employs the normalized Born approximation, or Born ratio, which divides measurements at the fluorescence wavelength by corresponding measurements at the excitation wavelength [22]. This approach allows direct computation of fluorescence parameters, without the intermediate determination of tissue optical properties [23]. Additionally, the Born ratio has been shown to grant a significant degree of invariance to inhomogeneities in the background optical properties of the medium. That is, the Born ratio corrects for differences between the modeled optical parameters and those present in vivo. Because of this correction, it is unclear whether the use of

more elaborate forward models is necessary, or if forward model simplifications will be sufficient to obtain optimal results.

Additionally, the use of prior information in the inverse problem also yields significant performance improvements on the resulting images. This raises questions regarding the interaction between structural prior knowledge obtained from XCT images in the forward and inverse problems, and to what degree inverse structural priors can compensate for simplifications of the forward model. To this end, we investigated inverse solutions both with and without structural priors, for each potential forward model, and examine the relative cost incurred by each subsequent model simplification.

In stand-alone FMT, where only the air-tissue surface may be known using surface extraction techniques, an internally homogeneous medium can be assumed with each optical property constant throughout the volume, and data normalization can be employed to correct for the effects of tissue optical heterogeneity [65]. For imaging of whole animals, the resulting parameter values are commonly set in the range of $\mu_a = 0.3cm^{-1}$ and $\mu'_s = 10cm^{-1}$, to correspond with the average values of bulk soft tissue. In contrast, the X-ray CT component of the hybrid approach provides knowledge of the internal geometry which can enable the implementation of more elaborate photon propagation models to account for the differences in absorption and scattering present within each organ or tissue. This improved modeling yields sensitivity functions which more accurately reflect the physical diffusion taking place within the animal. However, the introduction of additional optical parameters complicates the implementation of these models. Anatomic images are frequently segmented into a number of discrete regions corresponding to individual organs or optically similar tissue types, each of which must be assigned absorption and scattering values which accurately reflect those present in vivo. One option for selecting these values is to explicitly calculate them through solution of the diffuse optical tomography (DOT) problem [21]. This approach is theoretically the most accurate, but leads to increased complexity and additional computation requirements. Importantly,

this calculation may include and therefore propagate errors to the fluorescence reconstruction problem that can bias the final fluorescence image. An alternate approach that can be employed when constructing forward models for optical tomography is to employ average parameter values from ranges that exist or may be measured for the different tissue types segmented on the CT image. While perhaps not as accurate as explicit parameter estimation on a per animal bases, this approach is offset by the simplicity and lack of additional computation. What remains to be established is the tradeoff in terms of reconstruction accuracy that results from having a mismatch in modeled tissue optical parameters.

4.2 Methods

To examine the effects of parameter selection upon the resulting solutions, we assumed a three-dimensional geometry representative of the murine chest, as illustrated in Figure 4.1(a). This geometry was based on an *in vivo* X-ray CT of a nude mouse, segmented to differentiate heart, lung, liver, and bone from surrounding soft tissue. It is thus both physically realistic and representative of geometries for *in vivo* multimodal imaging of the chest cavity, a region that presents a difficult tomographic problem due to the high degree of optical inhomogeneity. Two hexahedral finite element method (FEM) meshes were generated, one with approximately 64k nodes, used for data generation, and a more coarse 11k node mesh for data inversion.

Optical parameter ranges for each tissue were obtained from existing literature (Tables 4.1 and 4.2). From these ranges, values were selected to create four models of light diffusion. We will refer to these models as the “matched,” “mismatched,” “mid-range,” and homogeneous models. The matched model is so-called because the high resolution model used for data generation employed the same set of tissue optical parameters. For each tissue type, the optical parameter value in the matched model was selected to be at one extreme of the associated published range. This allowed us to select optical

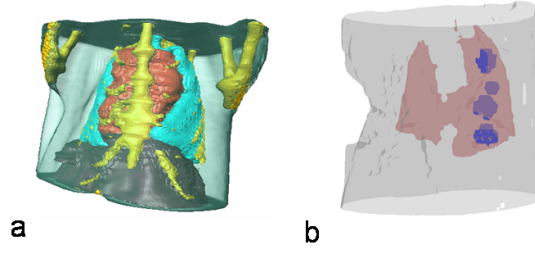


Figure 4.1: a) Rendering of FEM geometry b) Rendering showing exterior surface, lung surface, and location of all four fluorescing inclusions used in combination to generate simulated data measurements.

parameters for the mismatched model at the opposite end of each range from those selected for the matched model. Given the physically realistic ranges and the use of the matched model for data generation, this represents a worst case scenario, as the mismatch in μ_a and μ'_s was maximized with respect to values used for data generation. The mid-range model used values at the center of their corresponding ranges. Assuming that this represents the average absorption and scattering parameters for each tissue type, this choice will on average minimize the mismatch between the model and those parameters present *in vivo*. Finally, the homogeneous model assumed that the optical parameters were uniform throughout the medium, equivalent to assuming no prior knowledge about the animal's internal structure. These values are not shown in the tables, and were set to $\mu_a = 0.3cm^{-1}$ and $\mu'_s = 10cm^{-1}$ for all tissues.

Each of these four sets of model parameters was used to construct a diffusion problem on the anatomically defined finite element mesh. Using the Deal.II finite element

Table 4.1: Modeled and Published values for μ_a .

μ_a	Published Range	Model		
		Matched	Mismatched	Mid Range
Tissue		0.34	0.28	0.30
Bone		0.10	0.10	0.10
Lung	0.20 - 0.30	0.30	0.20	0.25
Heart	0.30 - 0.40	0.30	0.40	0.35
Liver	0.40 - 0.60	0.60	0.40	0.50

Table 4.2: Modeled and Published values for μ'_s

μ'_s	Published Range	Model		
		Matched	Mismatched	Mid Range
Tissue		12	10	10
Bone		20	17	20
Lung	25-35	35	25	30
Heart	20-25	20	25	23
Liver	10-15	15	10	13

libraries [?, ?], solutions to the diffusion approximation were obtained for point sources at each of source and detector locations. These solutions can be seen as Green's functions associated with the diffusing system, and can be used to construct the appropriate normalized Born models using the formula [22]:

$$w(r_s, r_d, r) = \frac{G(r_s, r)G(r_d, r)}{G(r_s, r_d)} \quad (4.1)$$

where $w(r_s, r_d, r)$ denotes the sensitivity of a measurement collected at point r_d , to fluorescence at a point r , given a point source of appropriate wavelength at point r_s . The function $G(r_1, r_2)$ denotes the solution to the diffusion problem at point r_2 , given a point source at r_1 .

Five fluorescing inclusions were constructed on this geometry, differing in location and physical dimensions. In each case, the boundaries of the inclusion were defined, and interior voxels given Gaussian distributed fluorescence intensities. The first simulates a situation where the fluorescent probe is spread throughout one entire lung, such as might be found when imaging lung inflammation. The remaining four targets were smaller, roughly spherical inclusions, as illustrated in 4.1(b). Three of these were approximately $3mm$ in diameter, while the fourth had a diameter of $2mm$. Each was located at a different, non-overlapping location within the lung, to cover a range of possible interference from other organs.

These inclusions were used to construct a total of 16 imaging scenarios. The first scenario consisted only of the full lung inclusion, while the remaining 15 data sets correspond to all possible combinations of 1-4 of the smaller inclusions. For each scenario,

multi-angle collection of diffuse data in a transmission geometry was assumed. At each collection angle, a 3×10 grid of source locations was defined, with overall dimensions $0.8\text{cm} \times 1.8\text{cm}$. A corresponding set of detectors locations was defined, using a 10×10 grid of size $1.6\text{cm} \times 1.0\text{cm}$. Using 17 projections spaced evenly every 20 degrees these values resulted in a total of 51000 source-detector pairs. Simulated data with 10% added shot noise was generated using the fine resolution mesh for Green's function computation, and the linear model presented above.

4.3 Results

Reconstructions for every data set-model combination were obtained by solving the Tikhonov regularized least squares problem:

$$\hat{x} = \arg \min_x \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2 \quad (4.2)$$

with the regularization parameter λ selected to minimize 2-norm error with the known true image. Solution of the above minimization was implemented using 50 iterations of the LSQR algorithm [75]. Each data set was reconstructed twice; once using the full weight matrix, and once using a structural prior model which constrained image values to lie solely within the lung region. The prior model was implemented by eliminating the elements of x and corresponding columns of A associated with voxels lying outside the lung region. For display purposes, the values of these voxels were then set to zero.

For each solution, 2-norm error with respect to ground truth was computed as:

$$e_i = \frac{\|x_i - x_{true}\|_2^2}{\|x_{true}\|_2^2}. \quad (4.3)$$

These errors are plotted for the the standard and *a priori* inversion techniques in Figures 4.2(a) and 4.2(b), respectively. In both cases, the matched model consistency provides the lowest reconstruction error, with the mid-range model yielding the second lowest error. For the standard reconstruction, the mismatched model also consistently outperforms the homogeneous model. When using the simple *a priori* approach, the errors obtained with

each model are significantly lower than those seen with the standard solution method. However, unlike the standard approach, there are several cases, when imaging three or four of the small inclusions, where the homogeneous model yields a lower 2-norm error than the mismatched model. Additionally, the spread in errors between the different models has increased.

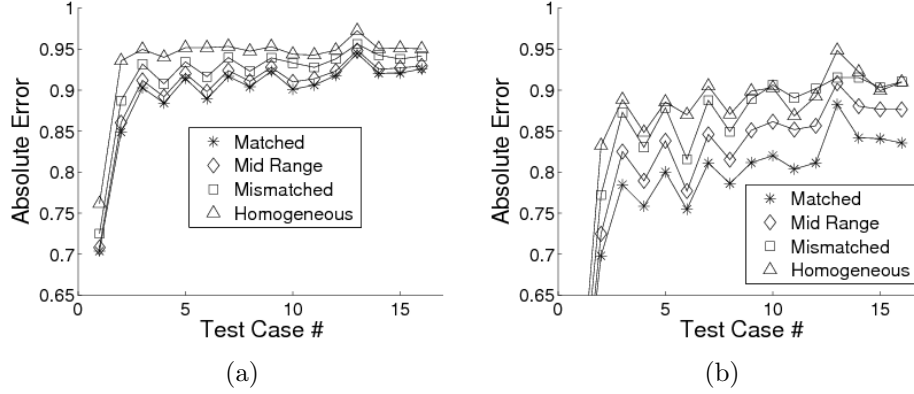


Figure 4.2: **Absolute 2-norm Error** using a) standard and b) *a priori* inversion. Results show consistent improvement when using mid-range and matched parameter models. Error levels for the full lung inclusion incorporating *a priori* structure in the inverse problem are not shown to better display other results.

To give a better comparison of the relative performance gain achieved by moving to a more complex model, we also computed the errors relative to the error seen with the matched forward model:

$$e_{i,relative} = \frac{e_i}{e_{matched}}. \quad (4.4)$$

These values are plotted in Figures 4.3(a) and 4.3(b). As also seen in the absolute error plots, the homogeneous model consistently shows the largest errors when using the standard inversion approach. As more accurate model parameters are incorporated, solution error correspondingly decreases in all cases. The mismatched model offers some improvement over the homogeneous, while the mid-range model consistently shows less than a 1% increase in error. Average relative error increases of 0.74%, 2.5%, and 4.8% are seen for the mid-range, mismatched and homogeneous models, respectively, as compared

to the matched model.

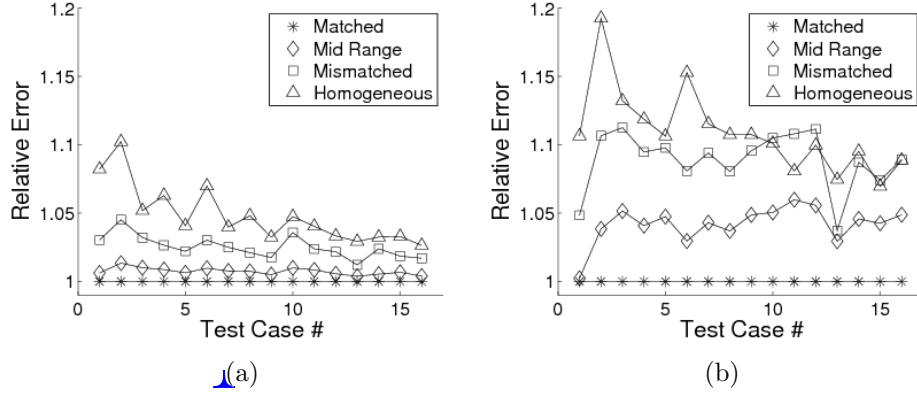


Figure 4.3: **Relative 2-norm Error** using a) standard and b) apriori inversion. Note that relative penalty for using the incorrect model is significantly higher when using *apriori* inversion techniques.

The situation changes slightly when incorporating structural prior knowledge into the inversion process. The full fluorescing lung (Test Case #1) is reconstructed with the lowest overall error levels, as the structure of the image is provided by the prior information. When reconstructing the smaller targets, the homogeneous model yields the highest error in 12 out of 16 trials, with the mismatched model yielding the highest error for the remaining 4 cases. Interestingly, while the use of prior information offered consistently lower error levels, the relative penalty for using an incorrect model increased. The average relative increase in error was 4.2% for the mid-range, 8.9% for the mismatched, and 10.9% for the fully homogeneous model, significantly higher than was seen without prior knowledge. These values reflect an increase in both the absolute error resulting from improper model selection, as well as the proportion of total error. This indicates that while structural *a priori* information in the inverse problem alone can offer improved performance, the biggest gains are to be made by a combination of prior knowledge and improved diffusion modeling.

Sample reconstruction slices from the full lung target using apriori knowledge are shown in Figure 4.4. While all models obtain the true structure of the lung, the matched

and mid-range models more accurately resolve the quantitative values. When reconstructing smaller inclusions, the more accurate models consistently offered improved separation of objects and lower 2-norm errors than the homogeneous and mismatched parameter models.

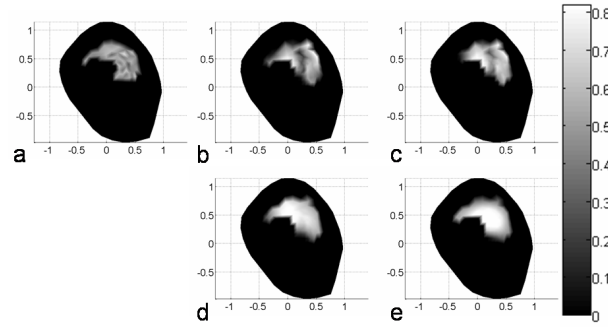


Figure 4.4: a) True full-lung image. b-e) Reconstructions of inclusion using: b) Matched model c) Mismatched Model d) Mid-range model e) Homogeneous Model

Our results suggest that for imaging within the murine chest, the use of established ranges for tissue optical parameters does not significantly degrade reconstruction quality, even when mismatch in values remains. Furthermore, by using values from the center of the established ranges, results can be obtained which consistently offer less than a 5% increase in error as compared to using perfectly matched parameters. In contrast, the use of homogeneous models results in significantly higher error levels and reduced image fidelity. We conclude that structural CT data can provide improvements to diffusion modeling for fluorescence molecular tomography without the additional experimental and computational complications of explicit parameter estimation.

4.4 Acknowledgments

The work in this chapter was completed in conjunction with Ralf Schulz at the Institute for Biological and Medical Imaging, Helmholtz Center Munich and Technical University of Munich. He developed the FEM modeling tools used for this study, and his assistance in designing the experiments was highly appreciated.

Chapter 5

MultiModality Imaging

5.1 Introduction

While significant improvements have been made in stand-alone FMT imaging systems, one drawback of fluorescence (as well as endogenous) optical contrast imaging using diffusion tomography techniques is the inherently low spatial resolution of the resulting solutions. Smoothness in the kernel results in poor conditioning of the linearized forward model, ultimately limiting resolution to objects of approximately millimeter scale and larger [4]. Given this low spatial resolution, there has recently been increasing interest in multimodality imaging [27–31, 86–94]. These techniques seek to combine the spatial resolution achievable by CT and MRI methods with the functional information available from optical techniques to obtain improved localization and quantification of contrast agent concentrations. The structural information can be used to improve both the forward and inverse aspects of the tomography problem. Techniques such as the finite element method (FEM) allow for solution of the diffusion problem on complex inhomogeneous tissue geometries, thereby improving the accuracy of the associated physical model. Additionally, the anatomical structure can be used to construct a prior model of the fluorescence image to be reconstructed. When applied to the inverse problem of recovering an image of the fluorescence distribution from a set of collected surface intensity measurements, this prior information can guide the reconstruction process to yield images that are both qualitatively and quantitatively superior.

Several methods currently exist for constructing such prior models, all of which share similar characteristics [30, 30, 85, 86, 95]. Most typically begin by labeling voxels in the high resolution structural image according to tissue type. Registration between the modalities then allows the assignment of a tissue label to each voxel in the FMT reconstruction. This labeling divides the solution image into a number of segments, each comprising a group of voxels with a common tissue label. Image constraints are then established on each segment, based on the assumption that the anatomic and functional images share a common structure.

The simplest method by which spatial structural information can be used is to constrain the solution to certain regions of the volume. If it is known *a priori* that fluorescence (or other contrast agent) can only appear in specific areas, then allowing the solution to take nonzero values only within that region can be an effective way of incorporating that structural information. For example, in the case of functional diffuse optical imaging of the human brain, it is known that changes in hemoglobin and deoxyhemoglobin concentrations are going to be occurring within the cortex of the brain, but not within the skull or cerebrospinal fluid [90]. By only allowing the inversion algorithm to place perturbations only within the cortical region, the accuracy of the reconstructed images is increased.

An alternate method of using structural information, proposed by Brian Pogue's group at Dartmouth, makes the assumption that the distribution of contrast agent is well correlated with the physical structure of the medium [27, 28, 30, 31, 94]. Further assuming that the concentration of contrast agent will be relatively smooth across each physical domain, a regularization matrix, with structure similar to that of a Laplacian regularizer, can be constructed. Given a total of N physical regions $\{\Omega_1, \Omega_2, \dots, \Omega_N\}$,

each with number of elements n_i

$$\mathbf{L}_{ij} = \begin{cases} 1 & i = j \\ \frac{1}{n_k} & x_i, x_j \in \Omega_k \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$

This encourages the solution within each spatial region to be uniform, while still allowing for sharp edges at the boundaries of each region.

A third proposed solution uses what the authors refer to as a hierarchical Bayesian approach to solving a maximum a posteriori (MAP) statistical problem [86]. Their overall problem takes the form of a MAP problem, using the joint probability density:

$$\begin{aligned} p(y, x, \lambda, \mu, \sigma | C) &= p(y, \lambda | x) p(x | \mu, \sigma, C) p(\mu | C) p(\sigma | C) \\ &= \frac{1}{K(2\pi)^{3N/2} |\Lambda_\zeta(\lambda)|^{1/2} |\Lambda_x(\sigma)|^{1/2} |\Lambda_\mu(\vartheta)|^{1/2} |\Lambda_\sigma(\gamma)|^{1/2}} \\ &\times \exp \left[-\frac{1}{2} \left(\|y - Wx\|_{\Lambda_\zeta^{-1}(\lambda)}^2 + \|x - \mu\|_{\Lambda_x^{-1}(\sigma)}^2 + \right. \right. \\ &\quad \left. \left. \|\mu - \bar{\mu}\|_{\Lambda_\mu^{-1}(\vartheta)}^2 + \|\sigma - \mu_\sigma\|_{\Lambda_\sigma^{-1}(\gamma)}^2 \right) \right] \end{aligned} \quad (5.2)$$

Here, $\Lambda_\zeta^{-1}(\lambda)$, $\Lambda_x^{-1}(\sigma)$, $\Lambda_\mu^{-1}(\vartheta)$, and $\Lambda_\sigma^{-1}(\gamma)$ are the measurement, image, image mean, and image covariance covariance matrices, respectively. The variables μ , $\bar{\mu}$, and μ_σ represent the mean of the image, the mean of the image mean, and the mean of the image covariance, respectively. The final two, $\bar{\mu}$ and μ_σ , are set based on analysis of the literature. The remaining values are left as hyperparameters, to construct a cost function:

$$\Psi(x, \lambda, \mu, \sigma) = -\log p(y, x, \lambda, \mu, \sigma | C) \quad (5.3)$$

With C being the particular image segmentation. Solution of this problem is done by repeatedly alternating between single conjugate gradient iterations to update the value of x , and individual MAP estimation of each of the hyperparameters.

In this section, we explore two aspects of the multimodality problem not directly considered in earlier reports. First, simply as a result of the imaging physics, multimodal imaging approaches generally use a high resolution structural image to enhance

low resolution images of functional activity. In the case of FMT-CT imaging, this means that each voxel in the low resolution FMT reconstruction will cover several voxels in the high resolution CT structural image. This inherent difference in spatial resolution has not, however, generally been reflected in the formation of the FMT inverse problem. While each CT voxel is assigned a single anatomic type during the initial segmentation, FMT voxels near tissue boundaries will potentially span two or more tissue types. By associating each FMT voxel with only a single tissue type, errors will result in tissue labeling, likely causing subsequent artifacts in the reconstruction. We note here that we use the term voxel to identify individual elements of the CT or FMT images, while segment refers to a group of voxels that share a common tissue type labeling at either the CT or FMT image scales.

Second, existing methods typically use *a priori* assumptions which are applied in an identical manner to each data set. Smoothness constraints are applied uniformly to all segments, and prior estimates of signal intensity are likewise fixed for all data sets. However, each segment of each experimental volume may have unique intensity and smoothness levels, which should be taken into account if they are to be constrained by the prior model. These image features can often be defined using a small number of parameters, which, if estimated from the collected data, could provide customized prior models offering improved reconstruction results.

This type of customized regularization is potentially of great benefit for imaging fluorescent targets. While endogenous optical contrast is often well correlated with physical structure, the fluorescent probes imaged by FMT systems are specific to molecular activity levels which may not correlate perfectly with anatomy, and whose intensity levels will not be known *a priori*. Accurately identifying regions of activity and subsequent selection of appropriate regularization levels would enable segments such as the background to be highly constrained, while permitting the remaining voxels in segments of higher activity to best account for the collected data.

To those ends, we introduce here a new space varying regularization technique for the incorporation of CT or other structural information into the solution of inverse diffusion problems. Our method addresses the fundamental resolution differences by labeling each FMT solution voxel as a linear mixture of the tissue types identified in the high resolution CT image. At the interface between physical regions, the resulting prior model is then able to incorporate information from multiple regions, more accurately representing the underlying anatomy.

Additionally, the segmental labeling is used to create a low dimensional inverse problem that yields a single intensity value for each anatomic segment. We then use these solution values as parameters to define a spatially varying regularization term for the full resolution FMT problem. The central idea of this paper is that the solution of the parameterized inverse allows us to construct customized constraints that vary depending on the collected data and spatial segmentation. The values obtained from the low dimensional problem provide an initial estimate of the total fluorescence activity present within each of the regions, and thus we have information about where we expect the highest intensities to lie.

We present two approaches for incorporating this information into the inverse problem as a spatially varying regularization term. In both cases, we use the solution of the parameterized problem to define a separate regularization term for each physical segment. This is then used in conjunction with the partial labeling map to assign a particular regularization level to each individual FMT voxel based on the underlying tissue types. The difference between the approaches lies in how the spatially varying mean values for the FMT voxels are applied. The first method uses the parameterized solution and partial labeling matrix to apply a spatially varying mean value prior, while the second forgoes this step. In both cases, the degree of regularization applied within each region is dependent upon the the solution of the low dimensional parameterized problem. Specifically, regions corresponding to a low value in the parameterized solution are regularized

more than regions with a high solution value. The assumption is that regions associated with low parameterized solution values are background regions which should be tightly constrained to prevent imaging artifacts from arising. Similarly, high values of the parameterized solution are assumed to correspond to regions of high fluorescence activity, and are subsequently regularized less to allow for accurate reconstruction of fluorescence intensities. In this way, we establish an intermediate step between hard priors, which reconstruct values on only a limited number of voxels (for example, one value within each particular physical region), and naive approaches which regularize each voxel uniformly.

5.2 Structural Priors

Given a CT or other structural images properly registered to the FMT coordinate system, individual voxels within the FMT image can be labeled with one or more tissue types. The segments arising from this labeling allows construction of a low dimensional problem that produces as a result a single intensity value for each physical segment. These values can then be used to construct a dataset-specific spatially varying regularization term for use in the full resolution FMT problem.

5.2.1 Low-Dimensional Parameterized Inverse

In order to employ structural information in the inverse problem, voxels in the FMT solution space must first be associated with the structural image. Modalities such as X-ray CT, however, have significantly better physical resolutions than FMT. Each FMT solution voxel will therefore occupy the same space as several voxels in the structural image. In the vicinity of tissue boundaries this difference in resolution means that each FMT voxel will potentially include more than one anatomic region. We propose to use a partial volume approach that models each FMT voxel as a mixture of multiple segments. This allows for the anatomical structure to be correctly related to the inverse problem.

To begin, we assume the existence of a fluorescence image at the CT image scale,

denoted \mathbf{x}_{CT} . This can then be related to the image at the FMT image scale as:

$$\mathbf{x}_{FMT} = \mathbf{C}_{CT}^{FMT} \mathbf{x}_{CT}, \quad (5.4)$$

where the matrix \mathbf{x}_{CT} , size $N_{FMTvoxels} \times N_{CTvoxels}$, computes the value of each FMT voxel as the average of the underlying CT voxels. We then introduce a second matrix $\mathbf{C}_{Segments}^{CT}$, size $N_{CTvoxels} \times N_{Segments}$, which maps each CT voxel to its underlying tissue type. The product of these two matrices: $\mathbf{C} = \mathbf{C}_{CT}^{FMT} \mathbf{C}_{Segments}^{CT}$, size $N_{FMTvoxels} \times N_{segments}$, will thus allow each FMT voxel to be fractionally labeled with multiple tissue types, depending on the labels of the underlying CT voxels. The individual elements \mathbf{C}_{ij} will then be the fraction of voxel i which lies within segment j . This process is depicted pictorially in Figure 5.1.

Using this matrix, a parameterized version of the inverse problem in (2.4) can be written as:

$$\hat{\mu}_x = \arg \min_{\mu_x} \|\mathbf{Q}(\mathbf{W}\mathbf{C}\mu_x - u)\|_2^2 = \arg \min_{\mu_x} \|\mathbf{Q}(\tilde{\mathbf{W}}\mu_x - u)\|_2^2 \quad (5.5)$$

where $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{C}$, of size $N_{data} \times N_{Segments}$ and μ_x is a vector of length $N_{Segments}$. This formulation is equivalent to obtaining an inverse solution using a piecewise constant basis defined by the anatomical segmentation. Alternately, it can be seen as a method for estimating a mean fluorescence intensity for each tissue type or anatomical region. Conceptually, the idea is that by solving a reduced dimensional problem using the structurally defined basis, we can use the collected data to provide information about the fluorescence content of each region. We will then use this reconstruction to build a regularization matrix customized to each specific multimodal data set.

Because negative fluorescence concentrations are non-physical, we would like to constrain the solution of (5.5) to be non-negative. The low dimensionality of the parameterized problem allows us to use the modified residual norm steepest descent (MRNSD) approach without significantly increasing total computation time [96]. MRNSD requires a non-zero initialization, which we achieve using a normalized projection of the data onto

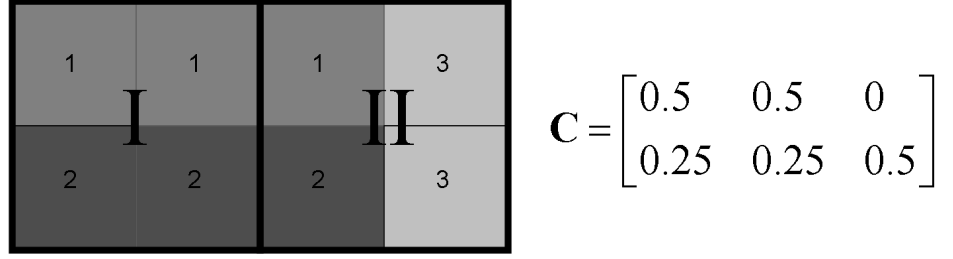


Figure 5.1: **Partial Volume Labeling** Construction of partial volume labeling matrix. Here, voxel I lies 50% within region 1 and 50% within region 2, while voxel II lies 25% within region 1, 25% within region 2, and 50% within region 3. This is reflected in the rows of the matrix C , each of which is of unit 1-norm.

the segments:

$$\hat{\mu}_{xj}^{(0)} = \frac{1}{N_{Segments}} \frac{u^T \mathbf{W}_{*j}}{\|\mathbf{W}_{*j}\|_2} \quad (5.6)$$

with $\tilde{\mathbf{W}}_{*j}$ denoting the j th column of the matrix \tilde{W} . The quantity $\hat{\mu}_{xj}^{(i)}$ corresponds to the value associated with the j^{th} region at iteration i .

5.2.2 Regularization Construction

Solution of (5.5) then yields a single value for each anatomic segment, which we will use in the construction of an appropriate spatially varying regularization term. If, as suggested earlier, these values are viewed as estimates of the mean for each segment, or alternately as an initial estimate of the image [78], then (2.4) could be modified to take the form:

$$\hat{\mathbf{x}} = \arg \min_x \|\mathbf{Q}(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2 + \lambda^2 \|\mathbf{L}(\mathbf{x} - \mathbf{C}\mu_x)\|_2^2 \quad (5.7)$$

which explicitly applies the μ_x values as spatially varying prior or initial estimates of mean intensity over the segments. This formulation also has a statistical interpretation as the maximum a posteriori (MAP) estimator of \mathbf{x} , given the collected data u , noise covariance $(\mathbf{Q}^T \mathbf{Q})^{-1}$, and the Gaussian image prior $x \sim N(\mathbf{C}\mu_x, \lambda^{-2}(\mathbf{L}^T \mathbf{L})^{-1})$.

Though optical contrast is often correlated with physical structure, the fluorescent probes that serve as the contrast agents for FMT imaging do not appear on a CT scan.

Moreover, their ability to traverse anatomical boundaries means it is possible for probe to be located within only a small portion of a physical segment, be present within more than one segment, or even be contiguous across an anatomical border. Thus, the assumption of a uniform mean across the entire segment may not strictly hold for fluorescence imaging. But as we will show in the results section, improved reconstructions can be obtained when this assumption is not explicitly satisfied. However, owing to these possibilities, we also consider the original minimization equation:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{Q}(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2 + \lambda^2 \|\mathbf{L}\mathbf{x}\|_2^2 \quad (5.8)$$

which does not explicitly impose mean values upon the solutions in each region, and can be interpreted as a MAP solution given the prior model $\mathbf{x} \sim N(0, \lambda^{-2}(\mathbf{L}^T \mathbf{L})^{-1})$.

In both cases, the matrix \mathbf{L} is constructed as a diagonal matrix, with spatially varying diagonal elements derived from the low dimensional solution:

$$\mathbf{L} = \text{diag}(\mathbf{C}\alpha). \quad (5.9)$$

We define $\alpha = f(\mu_x)$ for some function f , with \mathbf{C} the same matrix used in (5.5) to build the parameterized problem. Thus a regularization level is defined for each anatomic segment, and individualized regularization for each FMT voxel is then generated as a mixture of those intensities using the matrix \mathbf{C} . The function $f(\mu_x)$ can thus be seen in the statistical interpretation as generating the inverse of the standard deviation for each individual voxel. We want to select the function $f(\mu_x)$ such that an increase in mean value corresponds to a decrease in the corresponding regularization level. Regions with a low parameterized solution value will thus be treated as background regions, and more heavily regularized, while high parameterized solutions will lead to corresponding regions with lower levels of regularization.

A straightforward choice for $f(\mu_x)$ which achieves this goal corresponds to a model where the variance is proportional to the mean, as in the Gaussian approximation to a

Poisson process, resulting in the relationship:

$$\alpha_i = \frac{1}{\sqrt{\mu_{xi}}} \quad (5.10)$$

Unfortunately, this choice for α_i is unstable for small values of μ_{xi} , which are expected to appear frequently as a result of the non-negativity constraint applied by the MRNSD algorithm. In place of the above equation, we use the modified version:

$$\alpha_i = \sqrt{\frac{(1 + \beta) \max(\mu_{xj} | \forall j)}{\mu_{xi} + \beta \max(\mu_{xj} | \forall j)}}, \quad (5.11)$$

which constrains α_i to be in the range $\alpha_i \in [1, \sqrt{1 + 1/\beta}]$. This alteration prevents regions from being infinitely regularized as μ_{xi} goes to zero, which would be equivalent to applying a hard prior and not reconstructing values within that region. Our approach avoids this and allows fluorescence intensities to be reconstructed at any location within the volume, but more heavily penalizes their appearance within background regions.

5.2.3 Summary of Approach

Given an appropriately segmented anatomical image, coregistered to the solution image, our approach can be summarized as follows:

- (1) Use the segmentation to assign partial volume labels to each of the solution voxels, and construct the matrix C .
- (2) Solve (5.5) using MRNSD to obtain a single parameter value for each region.
- (3) Use the solution to compute the values of the α_i 's using (5.11).
- (4) Construct the regularization matrix L using (5.9), and proceed to solve either (5.7) or (5.8).

5.3 Experimental Methods

To evaluate our spatially varying regularization technique, we investigated a range of simulated and experimental data sets imaging amyloid- β plaques in the brains of

transgenic mice exhibiting the symptoms of Alzheimer’s disease. These plaques exist only within the brain, and are known to preferentially form within the cortical region. This yields a challenging tomographic problem, because the fluorescent targets are distributed throughout the volume, and not contained to a single small region.

External geometries were obtained directly from the FMT system using silhouette images and a previously described volume carving algorithm [65], while all anatomic information was provided by a single CT data set, collected from a wild type mouse on an X-SPECT small animal imaging system (Gamma Medica, Northridge CA). Given the relatively fixed internal skull geometry, an affine transform was used to coregister the CT data set with the computed exterior surfaces using the eyes, teeth and base of skull as fiducial markers. Semi-automatic segmentation was achieved using the software package AMIRA (Visage Imaging, Carlsbad CA). Because of the preferential formation of plaques, the cortical region and remaining brain tissue were assigned as separate segments. This differentiation was used as prior knowledge in the inverse problem, while the forward model treated the entire brain as a single optically homogeneous region. Finite element method (FEM) solutions to the diffusion approximation were obtained using COMSOL (COMSOL Inc, Burlington MA) to construct the linear model for simulated data generation and all inversions.

5.3.1 Anatomic Subsegmentation

In initial experiments, large changes in the parameterized solution occurred with very small changes in the data or initialization, indicating poor conditioning of the reduced dimensional system. Further examination revealed that certain anatomical configurations result in a high degree of linear dependence among the columns of \tilde{W} (The matrix of size $N_{data} \times N_{segments}$ resulting from the aggregation of voxels belonging to each segment). The columns of \tilde{W} associated with the cortical and remaining brain tissue segments were highly correlated. This made it extremely difficult for the MRNSD

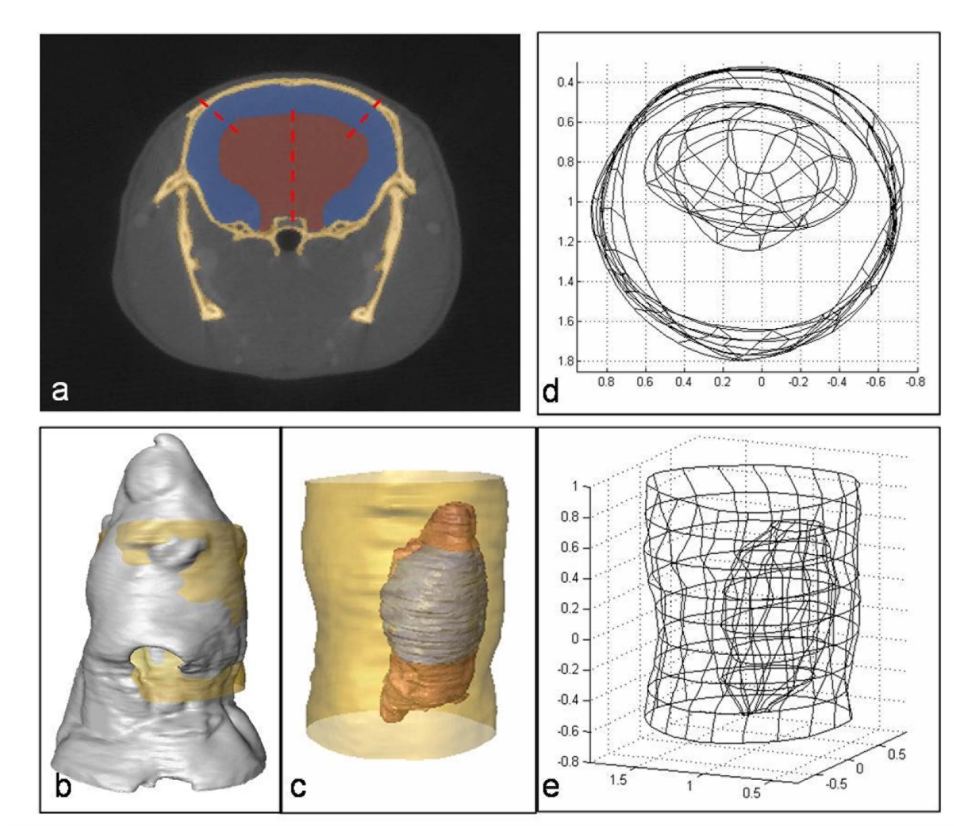


Figure 5.2: **CT-FMT Registration and Modeling Geometry** a) A 2D slice of the CT data, showing labeled brain (red) and cortical (blue) regions. Dotted red lines indicate where regions were subdivided prior to the inversion process. b) An isosurface from the CT image (gray) superimposed with surface computed by FMT volume carving algorithm (yellow). This alignment allows for use of CT information in the FMT problem. c) Alignment between modalities allows the brain (red) and cortical (blue) regions to be appropriately located within the surface computed by the FMT system (yellow). d-e) Contour information was extracted from the segmentation information and used to construct surface meshes prior to generation of a tetrahedral mesh for obtaining solutions to the diffusion equation using the finite element method (FEM).

algorithm to discriminate between segments, leaving the reconstructed parameters highly dependent upon initialization of the algorithm.

To correct this problem, we investigated the effects of subdividing the brain and cortex into smaller subsegments. For the work presented here, this was done in a purely geometrical manner, without taking the physical model or other attributes into account. We considered subdivision of both the central brain and cortical segments. The central brain was either left as a single segment, or divided into left and right hemispheres. To subdivide the cortex, the segment's centroid was located, and used as the reference point for dividing the cortex into $N \in \{1, 2, \dots, 10\}$ subsegments, using an equal central angle subdivision scheme. This process is illustrated in Figure 5.2.

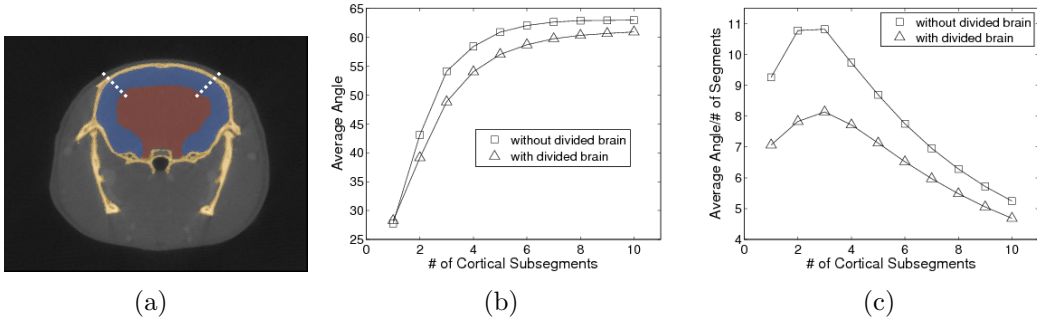


Figure 5.3: **Anatomic segmentation and subsegmentation.** a) The original segmentation separated the brain and brain cortex segments (red and blue segments, respectively), from the remaining soft tissue. Subsegmentation of the cortex divided it into three equal angle subsegments, as denoted by the white dashed lines. b) Plot of number of cortical subsegments against resulting average angle between segments, with (triangle) and without (square) subdivision of the central brain segment. c) Average angle divided by total number of subsegments plotted against number of cortical subsegments. Maximum is seen when cortex is divided into three subsegments.

For each configuration of subdivisions, the reduced dimensional matrix \tilde{W} was generated for analysis. To analyze the matrix, the angle between each pair of columns was computed as [70]:

$$\cos(\theta_{ij}) = \frac{\tilde{W}_{*i}^T \tilde{W}_{*j}}{\|\tilde{W}_{*i}\| \|\tilde{W}_{*j}\|} \quad (5.12)$$

The average of these angles was taken across all pairs of columns, to generate a single

average measure of the linear dependence present across columns of $\tilde{\mathbf{W}}$. As can be seen in Figure 5.3, this average angle rapidly increases with the number of cortical segments until it levels off at approximately 63 degrees by the time 7 segments are used. This suggests that additional subdivision of the cortex should help stabilize the reconstructions, but that there is a diminishing return from each additional subsegment. Interestingly, introducing subdivision of the central brain (indicated by the triangles in Fig 5.3) reduces the average angle by approximately 4 degrees as compared to leaving it in undivided.

Our goal is to minimize the number of additional subsegments introduced, thus preserving as much of the initial segmentation as practical. As a metric to measure the increased complexity against the benefits of increased average angle, we divided the average angle by the total number of subsegments used. As shown in Figure 5.3b, this metric has a maximum when the cortex is divided into three subsegments, and the remainder of the brain remains whole. Thus we employed this subdivision scheme for all reconstructions presented below.

5.3.2 Mouse Brain Simulation Studies

A series of simulation studies were generated to evaluate algorithmic performance in a range of situations. These scenarios were constructed to emulate scenarios expected to be seen in-vivo, as well as more artificial scenarios to test algorithm performance. The first test case is a simple arrangement, with each voxel more than 50% within the cortex assigned an intensity as a sample of a single non-zero mean Gaussian process.

The second test case uses the same cortex voxels, but applies a spatially varying mean value to the Gaussian process. The maximum is at the centerline of the head, and decreases linearly as it moves away to the left and right. Test case three is the same as test case two, with the addition of low level fluorescence within the remaining brain tissue. Test case number four has two small fluorescing regions within the cortex, while cases five and six have a spatially varying intensity within a portion of the cortex that

overlaps two subsegments, without and with the addition of background fluorescence elsewhere.

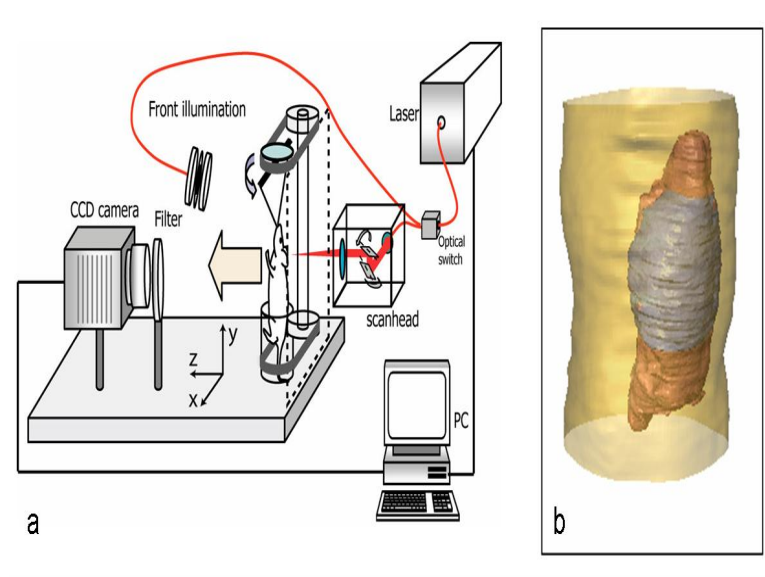


Figure 5.4: **System Diagram and Internal Head Geometry** a) Noncontact FMT imaging system diagram. b) Surface reconstructed with FMT system (yellow) overlaid with brain (red) and cortical (blue) segments obtained from CT structural image.

5.3.3 In Vivo Imaging Study

Longer human lifespans bring with them an increased prevalence of age related conditions such as neurodegenerative diseases [97]. Among these, Alzheimer's disease (AD) is the primary cause of neuronal degradation, accounting for between 42 and 81% of dementia cases. From an average onset of 80 years, AD incidence increases exponentially, with more than 50 percent of 90-95 year old individuals likely being symptomatic [98]. Furthermore, most similarly aged asymptomatic individuals are likely to acquire at least some amyloid- β plaques and hyperphosphorylated tau (tangles) which are the primary pathological hallmarks of the disease [99, 100]. While the specific importance of the neurofibrillary tangles has begun to attract significant interest in recent years [101, 102], the amyloid- β plaques have long been the primary target.

The widespread prevalence and significant impact of Alzheimer's disease upon both

patients and families have spurred the development of investigational tools for *in vivo* evaluation of both disease progression and the efficacy of potential therapeutic strategies. In the laboratory, several lines of transgenic animals exhibiting altered expression levels of amyloid precursor protein (APP) have been developed using mutations similar to those seen in early-onset familial AD (FAD) [103–106]. In particular, the APP23 line contains a mutated human APP gene that results in a 7-fold over-expression of APP compared to control animals [106], leading to the formation of amyloid plaques with age correlated increasing proliferation. Additionally, these mice show all other pathological hallmarks of Alzheimer’s disease such as hyperphosphorylated tau and manifest learning and memory deficits. Since their development, these animals have been widely used in research ranging from disease pathogenesis [107] to behavioral research [108] and imaging of altered brain vasculature [109–111].

Linked to the development of appropriate animal models for understanding and finding treatments for AD is the search for appropriate methods that can non-invasively image AD biomarkers *in vivo*. *In vivo* longitudinal imaging offers real time evaluation of drug efficacy and has been linked to accelerating therapeutic discovery. In response, imaging techniques such as positron emission tomography (PET) and magnetic resonance imaging (MRI) have been applied to AD visualization. Fluorodeoxyglucose (FDG) PET studies have examined AD related aberrations in glucose metabolism [112], while other radiotracers such as FDDNP [113,114] and Pittsburgh Compound-B [115,116] can directly assess amyloid plaque burden and tau bundles. Structural alterations such as decreased gray matter volume [117] and altered vasculature [109] have been imaged using magnetic resonance. Developments in fluorescence probes have yielded the ability to target amyloid- β plaques using the AO1987 probe [118]. Given the choice, fluorescence methods have advantages for biological research as they offer versatility that overcomes several of the technical difficulties of other modalities such as the use of ionizing radiation and the requirement for dedicated cyclotron facilities to produce adept PET probes or challenges

in visualizing molecular function with MRI. However current optical methods offer inadequate imaging performance. Fluorescence microscopy lacks the ability to penetrate deeper than a few hundred microns, preventing non-invasive imaging of the brain. Planar fluorescence imaging, previously utilized [118], offers single projection viewing and compromised accuracy due to the strong non-linear dependence of fluorescence intensity on activity depth and tissue optical properties. Furthermore, current state-of-the-art fluorescence tomography methods, developed to improve on planar imaging by using theoretical models of photon propagation in tissues, do not have adequate resolving power to produce accurate brain images. This is a particular problem when determining distributed fluorescence activity, i.e. spatially extended fluorescence patterns, as expected in neurodegenerative disease. Indeed, we demonstrate here that conventional fluorescence molecular tomography (FMT) of distributed fluorescence activity in the animal head is particularly challenging due to the optically heterogeneous structures and highly curving boundary characteristic of this problem.

***In Vivo* Imaging**

Two hours prior to imaging, each mouse was injected with a 1mg/kg dose of AO1987 (Novartis Institutes for BioMedical Research Inc) by tail vein injection. Additionally, each animal's head was shaved and a depilatory cream applied to ensure good coupling between the laser and tissue. Immediately prior to imaging, animals were anesthetized by tail vein injection of ketamine (100 mg/kg) and xylazine (10 mg/kg). This was supplemented as necessary during imaging by additional intraperitoneally injected doses. Animals were mounted vertically within the chamber, with their abdomen and lower extremities restrained within a cylindrical tube. To properly orient and restrain the head, the animal's incisor teeth were fastened to the top of the rotation stage with suture thread. Imaging sessions lasted approximately 1 hour from initial anesthetization to completion and euthanasia by CO₂ inhalation.

Mean FMT Data Signal

Collected FMT data were used to generate images approximating those collected *in vivo* with uniform illumination planar transmission systems. A single projection was selected, corresponding to the CCD camera being positioned directly over the dorsal surface of the head. Laser source locations were on the ventral side. Images from fluorescence and excitation channels were summed across all source locations to yield single images for each channel. By dividing each pixel of the fluorescence image by the corresponding image in the excitation image, the normalized Born ratio was applied to correct for differences in excitation intensity. Mean normalized intensity over the brain region was then computed, for comparison with *in vivo* tomographic reconstructions and *ex vivo* planar imaging results.

Ex vivo Imaging

Immediately following *in vivo* FMT imaging, each animal was perfused intra-aortically with a 4% perfomaldehyde in phosphate buffered saline (PBS) solution to fix and solidify brain tissue. Brains were surgically extracted and imaged whole using a planar reflectance imaging system before being manually sliced into 8 axial sections, approximately 1.5mm thick, for further examination. Illumination for planar reflectance imaging was with the same 650nm continuous wave diode laser used in the FMT system. Beam expanders enabled a planar illumination field, rather than the point illumination used for tomographic imaging. Data were collected at excitation and emission wavelengths using the same wavelength selection filters as the FMT system, as well as under white room light. Post-processing using ImageJ and Matlab (The Mathworks, Natick MA) was used to identify the cortical region of each slice and compute the mean normalized fluorescence signal using the Born ratio.

Confocal Imaging

Fixed brain slices were mounted onto CoverWell & Secure Seal Imaging Chambers

(Electron Microscopy Sciences, Harfield, PA) with Crystal Mouse (BioMeda, Foster City CA). Photographs of each slice were taken with an Olympus C7070 Digital Camera mounted to an Olympus Model SZX12 Research Stereo Microscope. Regions of interest were selected on these images to guide region selection for confocal microscopy. Confocal microscopy data were collected with a Zeiss LSM 510 Meta confocal microscope with the two photon (NLO) feature from a subset of these regions. Confocal images were acquired using the 635 nm HeNe laser. The emitted light was collected through a 650-710 nm bandpass filter. A 10 layer stack was collected of each field then collapsed to one maximum projection image. Data were analyzed using the CRi Maestro (CRi, Woburn, MA) measurement software to select plaques (defined as pixel clusters above certain threshold values); the sum of the selected plaques (area in pixels) was then normalized within that image.

5.3.4 Acknowledgements

I would like to acknowledge the contributions of several individuals, without whom this *in vivo* imaging study could not have been completed. Firstly, Dr Thomas Krucker at the Novartis Institutes for Biomedical Research (NIBR), who kindly provided the probe transgenic animals used for the study. Dr Ruben DeKleine, in the course of a year in the lab at MGH, was responsible for the tail vein injections of both probe and anaesthesia, as well as the post-mortem excision and imaging of the brains. Sarah MacLaurin and Akos Szilvasi, both also at NIBR, were responsible for the confocal imaging of the brain slices.

5.4 Results

5.4.1 Partial Volume Labeling

Figure 5.5 shows results for a simulated data set with fluorescence present evenly throughout the cortical region. Figure 5.5(b,c) show reconstructions where each FMT

voxel was assigned a single tissue type, while Figs. 5.5(d,e) are reconstructions of the same dataset using partial volume labeling. Subfigure (b) and (d) were constructed by incorporating a mean value into the prior model, while (c) and (e) only specify prior knowledge about the covariance. The original image used to generate the data is shown in Fig. 5.5a. Without the use of partial labeling, the structure of the segmentation is clearly visible, with hard edges and right angle corners that are not present in the original image. By incorporating partial volume labeling, however, these artifacts can largely be eliminated, yielding reconstructions with a much more natural appearance. Additionally, as indicated by the overlaid numerical values, the use of partial volume labeling reduces the relative 2-norm error present in the solutions.

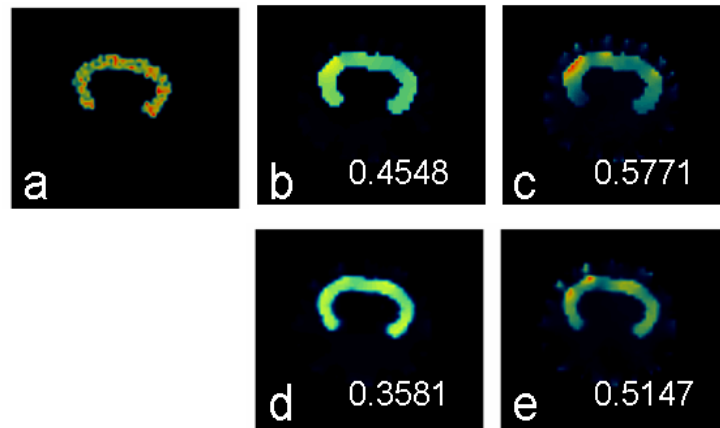


Figure 5.5: **Effects of Partial Volume Labeling:** a) Original Image b-c) Reconstructions without partial volume labeling d-e) Reconstructions employing partial volume labeling.

5.4.2 Simulations

Figure 5.6 shows inversion results for each of the simulated datasets. The original images are seen in the first column, while the second column shows reconstructions using standard Tikhonov regularization with the identity matrix. In all cases, these reconstructions fail to recover either the location or intensity of the fluorescence activity. Instead,

they are heavily corrupted by surface artifacts, and the more deeply situated fluorescence is blurred throughout the volume. Clearly, the identity matrix is an inappropriate regularization choice for imaging of the distributed phenomenon of interest.

The reconstructions in column 3 are obtained by applying a simple *a priori* technique which regularizes voxels outside the brain segments in the CT segmentation four times as heavily as those within. This has the effect of largely constraining the reconstruction to the brain segment by heavily penalizing solutions with fluorescence within the surrounding soft tissue. These reconstructions show a marked improvement over reconstructions using the identity matrix. The fluorescence is now inside the brain, but the different segments within the brain are not distinct, and the reconstructions still have a largely diffuse nature. For example, in test cases 5 and 6, the entire brain is resolved to have approximately the same fluorescence intensity, although the original images have higher intensity in the cortical segment.

Reconstructions using both region dependent means and variances are shown in the forth column. Here we see a much better differentiation of of the cortical region, as compared to the reconstructions where the images is simply restricted to lying within the brain, and a closer resemblance to the original images. Relative 2-norm error with respect to ground truth is also significantly reduced. However, these reconstructions yield highly uniform values on each of the regions, and fail to capture the shifting mean values seen in cases 5 and 6. These uniform values reflect the estimates of the mean value obtained in the first, low dimensional, inversion step. For datasets one, two, and three, where the voxels in each subsegment have similar intensities, the resulting reconstructions closely resemble ground truth. However, in test case six, applying the mean value actually performs worse with regards to 2-norm error than the simple *a priori* approach. Additionally, the reconstructions clearly show artifacts from the segmentation and subsegmentation procedures, suggesting that a more complex, data-driven subsegmentation procedure may be useful for providing further improvements in the reconstructions.

Finally, the fifth column shows reconstruction results using our region dependent variance model without the explicit use of mean values. As when implementing the mean values, the reconstructions seen are improvements over both the Tikhonov with the identity, and the simple *a priori* technique. In case 1, additional artifacts can be seen as compared to the image in column 4, and this is reflected in a higher 2-norm error. Reconstructions for cases 2-5 appear very similar to those obtained using the mean value, although the 2-norm error is higher. In test case 6, however, the reconstruction without the use of a mean value yields both the lowest 2-norm error, as well as the greatest subjective similarity to the original image.

5.4.3 In-Vivo FMT-CT Imaging of Amyloid- β Plaques

Six mice were imaged during the course of this study: four APP23 tg mice, at ages 17 to 28 months, and two APP23 wt control mice, ages 13 and 17 months. For all APP23 tg mice, the initial parameterized inversion identified one or more of the cortical segments as the region(s) most likely to contain fluorescence, a conclusion consistent with prior expectations. By contrast, applying an initial inversion step to the wt mouse shown in Figure 5.9 did not identify strong fluorescence within any individual region. This resulted in a reconstruction with a less structured fluorescence distribution and lower overall intensity, suggesting that activity may be the result of free fluorochrome or non-specific binding. Full resolution reconstructions of the APP23 tg mice in our algorithm's second step resulted in images showing fluorescence activity primarily within the cortical region of the brain.

Images constructed using our structurally guided multimodal algorithm consistently showed a marked improvement in quality and significant decrease in artifacts compared to those regularized using standard Tikhonov regularization with the identity matrix. As in the simulated tests, without structural prior information reconstructions of *in vivo* data were dominated by diffuse fluorescent masses and surface artifacts, offering

little qualitative or quantitative information about plaque localization. The developed multi-modal algorithm eliminated these artifacts and enabled quantitative measurement of fluorochrome concentrations which were correlated with measurements gathered *ex vivo* by other imaging methods detailed below.

Correlative ex vivo Imaging Immediately following *in vivo* imaging, each animal was euthanized and its brain surgically excised for imaging with a planar fluorescence reflectance system. Individual image slices were analyzed to determine mean normalized fluorescence intensity levels within the manually segmented cortex. The resulting values were then plotted against the mean *in vivo* reconstructed cortical intensity from the FMT system (Fig. 5.8) yielding a highly linear relationship ($R^2=0.9587$) between the signal observed by *ex vivo* imaging and multi-modal FMT reconstructed intensities. By contrast, the mean FMT input signal from a single projection, which approximates planar transmission imaging, shows a markedly lower correlation with the *ex vivo* measurements ($R^2=0.7422$). While correlation between mean FMT input signal and reconstructed values is observed ($R^2=0.8712$), it is lower than that present between *ex vivo* data and reconstructed values. Additionally, higher input values do not always yield higher reconstructed intensities, indicating that the tomographic process is properly correcting for variations in geometry and fluorochrome localization within the head.

Subjective evaluation of the *ex vivo* images further confirms the relationship between underlying activity and the FMT reconstructions (Fig. 5.10). No lesions are seen in the C57BL/6 control mouse, while varying degrees of plaque density appear in the cortex of the APP23 tg mice. Fluorescence localization in the FMT images correlates well with these *ex vivo* findings. While the resolution of the FMT system cannot resolve the individual lesions seen in the planar fluorescence images, the pattern of plaque deposits in the planar images correlates well with that seen in the FMT reconstructions.

Further validation of both the AO1987 probe and the multimodal FMT-CT imag-

ing technique were provided by confocal microscopic imaging of the excised brain slices. Confocal images of cortical and hippocampal brain regions showed dense plaque accumulations in the APP23 tg mice, while the images from control mice were void of all staining (Fig. 5.10). Thick slices were analyzed to quantify the total area within cortex and hippocampus occupied by amyloid- β plaques. The average fractional area for five APP23 mice (average age 23.6 months) was 11.62% (SEM 0.22) while control wt mice lacked any measurable signal. This was in accordance to previously established plaque load in these mice⁸. Notably larger blood vessels in the cortical area stained positive.

5.5 Optimal Parameter Estimation

An area in which further research could improve these techniques is with a more statistically appropriate method for determining the means and variances associated with each of the individual physical regions of the solution image. In this chapter, mean values within each region were estimated as:

$$\hat{\boldsymbol{\mu}}_x = \arg \min_{\boldsymbol{\mu}_x} \|\mathbf{Q}(\mathbf{WC}\boldsymbol{\mu}_x - \mathbf{u})\|_2^2 \quad (5.13)$$

with variances estimated as an analytic function of the means. If each region were truly piecewise constant, this approach would be the maximum likelihood estimator of the means. However, fluorescence intensity within each region will have some variability about the mean value that should be accounted for in the estimation process. I therefore assume a general image prior of:

$$\mathbf{x} \sim N(\mathbf{C}\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x(\boldsymbol{\mu}_x)). \quad (5.14)$$

Note that the covariance matrix $\boldsymbol{\Sigma}_x$ is assumed to be dependent upon the vector of mean values $\boldsymbol{\mu}_x$. This is based on the assumed availability of an estimate of $\boldsymbol{\mu}_x$, which is then used to compute variances estimates. The specific details of the covariance model are unimportant for these derivations. Only the existence of the dependence is relevant. If the covariance is assumed to be known, with no dependence upon the mean values, the

fixed point iteration derived below can be eliminated, and the process reduced to a single step. The above model can now be used to return to the original linear model, and rewrite it as:

$$\mathbf{u} = \mathbf{W}(\mathbf{C}\boldsymbol{\mu}_x + \boldsymbol{\Delta}_x) + \mathbf{n} \quad (5.15)$$

where \mathbf{n} is the measurement noise, distributed as $\mathbf{n} \sim N(0, \boldsymbol{\Sigma}_n)$, and $\boldsymbol{\Delta}_x$ is the vector of variations about each voxel's mean value, with $\boldsymbol{\Delta}_x \sim N(0, \boldsymbol{\Sigma}_x)$. This can be further rearranged to write:

$$\mathbf{u} = \mathbf{W}\mathbf{C}\boldsymbol{\mu}_x + \mathbf{n}_{tot} \quad (5.16)$$

with, $\mathbf{n}_{tot} = \mathbf{W}\boldsymbol{\Delta}_x + \mathbf{n}$, $\mathbf{n}_{tot} \sim N(0, \mathbf{W}\boldsymbol{\Sigma}_x\mathbf{W}^T + \boldsymbol{\Sigma}_n)$. Here, the variation around the background, and its resulting effects upon the measured data, are treated as additional noise, rather than either being ignored or treated as deterministic. From this, it is straightforward to show that the maximum likelihood solution is:

$$\hat{\boldsymbol{\mu}}_x = \arg \min_{\boldsymbol{\mu}_x} (\mathbf{u} - \mathbf{W}\mathbf{C}\boldsymbol{\mu}_x)^T (\mathbf{W}\boldsymbol{\Sigma}_x\mathbf{W}^T + \boldsymbol{\Sigma}_n)^{-1} (\mathbf{u} - \mathbf{W}\mathbf{C}\boldsymbol{\mu}_x) + \log(|\mathbf{W}\boldsymbol{\Sigma}_x\mathbf{W}^T + \boldsymbol{\Sigma}_n|). \quad (5.17)$$

Here, the dependency of $\boldsymbol{\Sigma}_x$ upon $\boldsymbol{\mu}_x$ makes direct solution of this minimization difficult. The determinant in the second term will be computationally expensive to compute, making even direct gradient based methods difficult to implement. However, if $\boldsymbol{\Sigma}_x$ were known and fixed, the second term would be constant, and therefore eliminated from the minimization, and the first term would simply be a weighted least squares problem. One possible solution to this is to use a currently available estimate of $\boldsymbol{\mu}_x$ to compute a fixed estimate of $\boldsymbol{\Sigma}_x$, which allows easy solution of the above minimization, providing an updated estimate of $\boldsymbol{\mu}_x$. Repeating this procedure results in a fixed point iteration:

$$\hat{\boldsymbol{\mu}}_x^{(n+1)} = \arg \min_{\boldsymbol{\mu}_x} (\mathbf{u} - \mathbf{W}\mathbf{C}\boldsymbol{\mu}_x)^T (\mathbf{W}\boldsymbol{\Sigma}_x^{(n)}\mathbf{W}^T + \boldsymbol{\Sigma}_n)^{-1} (\mathbf{u} - \mathbf{W}\mathbf{C}\boldsymbol{\mu}_x) \quad (5.18)$$

with $\boldsymbol{\Sigma}_x^{(n)} = \boldsymbol{\Sigma}_x(\mathbf{x}^{(n)})$, which is the same as solving:

$$\hat{\boldsymbol{\mu}}_x^{(n+1)} = \arg \min_{\boldsymbol{\mu}_x} \|(\mathbf{W}\boldsymbol{\Sigma}_x^{(n)}\mathbf{W}^T + \boldsymbol{\Sigma}_n)^{-1/2}(\mathbf{u} - \mathbf{W}\mathbf{C}\boldsymbol{\mu}_x)\|_2^2, \quad (5.19)$$

and which has the solution:

$$\hat{\boldsymbol{\mu}}_x^{(n+1)} = (\mathbf{C}^T \mathbf{W}^T (\mathbf{W} \boldsymbol{\Sigma}_x^{(n)} \mathbf{W}^T + \boldsymbol{\Sigma}_n)^{-1} \mathbf{W} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W}^T \mathbf{u} \quad (5.20)$$

Now, however, the issue arises of how to compute either the inverse or inverse square root of the matrix $\mathbf{W} \boldsymbol{\Sigma}_x^{(n)} \mathbf{W}^T + \boldsymbol{\Sigma}_n$. Given the size and dense nature of this matrix, explicit computation of the matrix inverse will be computationally expensive. By making use of the standard or generalized singular value decompositions, an approximate inverse can be computed in an iterative fashion which will likely require less overall computation, and be able to execute in a smaller memory footprint.

First, the matrix to be inverted can be rewritten as:

$$\mathbf{W} \boldsymbol{\Sigma}_x^{(n)} \mathbf{W}^T + \boldsymbol{\Sigma}_n = \begin{bmatrix} \mathbf{W} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_x & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_n \end{bmatrix} \begin{bmatrix} \mathbf{W}^T \\ \mathbf{I} \end{bmatrix} \quad (5.21)$$

$$= \boldsymbol{\Sigma}_n^{1/2} \begin{bmatrix} \boldsymbol{\Sigma}_n^{-1/2} \mathbf{W} \boldsymbol{\Sigma}_x^{1/2} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_x^{1/2} \mathbf{W}^T \boldsymbol{\Sigma}_n^{-1/2} \\ \mathbf{I} \end{bmatrix} \boldsymbol{\Sigma}_n^{1/2} \quad (5.22)$$

if the GSVD of the matrix pair $\{\boldsymbol{\Sigma}_n^{-1/2} \mathbf{W}, \boldsymbol{\Sigma}_x^{1/2}\}$ is taken, you obtain:

$$\boldsymbol{\Sigma}_n^{-1/2} \mathbf{W} = \mathbf{U} \mathbf{S} \mathbf{X}^{-1} \quad (5.23)$$

$$\boldsymbol{\Sigma}_x^{-1/2} = \mathbf{V} \mathbf{M} \mathbf{X}^{-1}. \quad (5.24)$$

The second equation can also be rearranged to yield $\boldsymbol{\Sigma}_x^{1/2} = \mathbf{X} \mathbf{M}^{-1} \mathbf{V}^T$, from which you can get: $\boldsymbol{\Sigma}_n^{-1/2} \mathbf{W} \boldsymbol{\Sigma}_x^{1/2} = \mathbf{U} \mathbf{S} \mathbf{M}^{-1} \mathbf{V}^T$, which is a standard SVD. I present both the SVD and GSVD forms, because the ease with which the two can be implemented depends on whether $\boldsymbol{\Sigma}_x^{-1/2}$ or $\boldsymbol{\Sigma}_x^{1/2}$ is more readily available, and which imposes a higher cost for matrix-vector multiplications (One matrix is likely to be sparse while the other will be dense due to the inversion). In either case, using the identity $\boldsymbol{\Sigma}_n^{-1/2} \mathbf{W} \boldsymbol{\Sigma}_x^{1/2} = \mathbf{U} \mathbf{S} \mathbf{M}^{-1} \mathbf{V}^T$ in (5.21), yields:

$$\mathbf{W} \boldsymbol{\Sigma}_x \mathbf{W}^T + \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_n^{1/2} (\mathbf{U} (\mathbf{S}^2 \mathbf{M}^{-2} + \mathbf{I}) \mathbf{U}^T) \boldsymbol{\Sigma}_n^{1/2} \quad (5.25)$$

and additionally:

$$(\mathbf{W}\Sigma_X\mathbf{W}^T + \Sigma_n)^{-1} = \Sigma_n^{-1/2} (\mathbf{U}(\mathbf{S}^2\mathbf{M}^{-2} + \mathbf{I})^{-1}\mathbf{U}^T) \Sigma_n^{-1/2}. \quad (5.26)$$

This formulation can be used in conjunction with the identity $\mathbf{W} = \Sigma_n^{1/2}\mathbf{U}\mathbf{S}\mathbf{M}^{-1}\mathbf{V}^T\Sigma_x^{-1/2}$ to rewrite (5.20):

$$\hat{\boldsymbol{\mu}}_x^{(n+1)} = (\mathbf{C}^T\mathbf{W}^T(\mathbf{W}\Sigma_x^{(n)}\mathbf{W}^T + \Sigma_n)^{-1}\mathbf{W}\mathbf{C})^{-1} \mathbf{C}^T\mathbf{W}^T\mathbf{u} \quad (5.27)$$

$$= (\mathbf{C}^T\mathbf{W}^T(\Sigma_n^{-1/2} (\mathbf{U}(\mathbf{S}^2\mathbf{M}^{-2} + \mathbf{I})^{-1}\mathbf{U}^T) \Sigma_n^{-1/2})\mathbf{W}\mathbf{C})^{-1} \mathbf{C}^T\mathbf{W}^T\mathbf{u} \quad (5.28)$$

$$= (\mathbf{C}^T\Sigma_x^{-1/2}\mathbf{V}\mathbf{S}^2\mathbf{M}^{-2}(\mathbf{S}^2\mathbf{M}^{-2} + \mathbf{I})\mathbf{V}^T\Sigma_x^{-1/2}\mathbf{C})^{-1} \mathbf{C}^T\mathbf{W}^T\mathbf{u} \quad (5.29)$$

$$= \mathbf{C}^{-1}\Sigma_x^{1/2}\mathbf{V}\mathbf{S}^{-2}\mathbf{M}^2(\mathbf{S}^2\mathbf{M}^{-2} + \mathbf{I})^{-1}\mathbf{V}^T\Sigma_x^{1/2}\mathbf{C}^{-T}\mathbf{C}^T\mathbf{W}^T\mathbf{u} \quad (5.30)$$

$$= \mathbf{C}^{-1}\Sigma_x^{1/2}\mathbf{V}\mathbf{S}^{-2}\mathbf{M}^2(\mathbf{S}^2\mathbf{M}^{-2} + \mathbf{I})^{-1}\mathbf{V}^T\Sigma_x^{1/2}\mathbf{C}^{-T}\mathbf{C}^T\Sigma_x^{-1/2}\mathbf{V}\mathbf{S}\mathbf{M}^{-1}\mathbf{U}^T\Sigma_n^{1/2}\mathbf{u} \quad (5.31)$$

$$= \mathbf{C}^{-1}\Sigma_x^{1/2}\mathbf{V}\mathbf{S}^{-1}\mathbf{M}(\mathbf{S}^2\mathbf{M}^{-2} + \mathbf{I})^{-1}\mathbf{U}^T\Sigma_n^{1/2}\mathbf{u}. \quad (5.32)$$

This final equation gives a method for iteratively estimating the values of $\boldsymbol{\mu}_x$ from the collected data that explicitly incorporates the knowledge that the image to be reconstructed will have some variability within each region. This approach does, however, require significantly more computation than the more basic estimation scheme used in Section 5.2.2. With that scheme, the system matrix $\mathbf{W}\mathbf{C}$ was significantly reduced from the original \mathbf{W} , and computation could be done very rapidly. The approach presented above, however, operates at the full dimensionality of \mathbf{W} , and thus each step of the fixed point iteration will require approximately as much time as solving the fully voxelated problem $\mathbf{W}\mathbf{x} = \mathbf{u}$. The question remains whether or not the estimates obtained from this more complex system offer sufficient improvement over the more basic approach to warrant the additional computation.

5.6 Discussion and Conclusions

Non-invasive fluorescence tomography of laboratory animals can become a natural extension of *in vitro* fluorescence assays and microscopy studies given that high image

fidelity and quantification accuracy can be achieved *in vivo*. Fluorescence is an essential modality in biological research and the extension of fluorescence imaging resonates with the biology culture. In addition, fluorescence imaging offers attractive characteristics such as the ability to simultaneously differentiate multiple markers based on multi-spectral imaging [119], the use of non-ionizing radiation and the widespread use of fluorescent proteins [120], recently also extending to the near-infrared [121–123]. The use of appropriate scrambled or inactive probes emitting in a different wavelength compares to the active probe, can further generate imaging protocols where each mouse serves as its own control, for example by independently resolving probe binding vs. probe bio-distribution, leading to increased accuracy and minimization of the animals required to obtain statistically significant results. While *in vivo* imaging does not replace traditional *ex vivo* studies or intravital microscopy measurements based on established microscopy protocols, it can be used to provide volumetric data at the tissue level and accurately guide the selection of animals for entering established high resolution imaging protocols for deciphering events at the cellular and sub-cellular level.

For these reasons, several methodologies have been described for quantitative fluorescence imaging of tissue. However none of these methods have proven so far sufficient for *in vivo* imaging of neurodegenerative disease. The particular challenge associated with optical imaging of neurodegenerative diseases is that the disease is rarely confined within a small volume of the brain but it is more typical that a distributed spread pattern arises in the brain. Therefore, while localized cancer-related fluorescence activity has been resolved with conventional fluorescence tomography in the animal brain [124], this technology has difficulty accurately visualizing distributed fluorescence bio-distributions as shown here with relevant simulations and *in vivo*.

We have presented here two new methods for the inclusion of *a priori* structural information in the construction of inverse diffusion images via data-informed multimodality space-varying regularization. In contrast to existing techniques, this approach accurately

associates solution voxels with their underlying tissues types, and then regularizes voxels based on these associations rather than imposing a prior structure directly on the solution. We achieve this by using a multiple label map to define a low dimensional parameterized problem which yields information about the relative importance of each region to the overall reconstruction. These values are then used to construct a space-varying regularization term based on the tissue labeling of each voxel.

These methods are used to demonstrate that a hybrid method based on state-of-the-art FMT measurement combined with CT structural information enables the *in vivo* visualization of plaque distributions in a mouse model for Alzheimer’s disease. We show that the use of this technique is essential to produce highly accurate fluorescence images, in contrast to using standard tomographic methods. Simulation studies demonstrate that even with perfect model information, accurate reconstruction of spatially distributed fluorescent inclusions can be a difficult problem without prior knowledge of the underlying anatomical structure. The ill-posed nature of the problem bias regularized solutions towards smooth solutions which do not reflect the underlying structure of the fluorescence distribution. Without the use of structural priors to compensate for this bias, reconstruction of distributed fluorescence is extremely difficult. By employing CT structural information in the reconstruction process, we are able to recover spatial and quantitative information with a much higher degree of accuracy than previously possible.

This capability provides an important new tool for *in vivo* imaging of Alzheimer’s disease progression within a single mouse, and suggests applicability to imaging other neurodegenerative diseases in the murine head. The information obtained from such studies could provide real time evaluation of the efficacy of new anti-amyloid treatments and insights into Alzheimer’s disease pathology over time. By utilizing the same method at several spectral windows, with corresponding fluorochromes or expressed fluorescent proteins from transgenic animals emitting at different spectral bands, different disease biomarkers and signaling pathways or cellular migration can be visualized including phys-

iological and molecular markers.

An important aspect of the method developed is the restriction of the use of *a priori* information to the definition of physical regions in the head; i.e. without other restrictions on fluorescence bio-distribution. This allows the use of CT structural information in a manner which does not overly impose itself onto the fluorescence reconstruction. Instead, the collected FMT measurements themselves are allowed to determine the final output. This is important because operator bias and/or conceptions of a desired output can strongly impact the resulting reconstructions when using *a priori* information in a less flexible manner, especially in fluorescence mode where the underlying spatial distribution does not necessarily correlate with anatomic structure, and thus the information required to make such assumptions is unavailable.

One drawback seen in the images from several of the mice with heavier cortical amyloid- β plaque loads was a degree of spatially non-uniform sensitivity to the presence of fluorochrome. Lesions present in the dorsal portion of the cortex are more easily reconstructed than those present in the left and right lateral segments, and result in higher reconstructed intensities than the lateral segments. This can be explained by the geometry of the head, differences in absorption between the brain and surrounding tissue, and the presence of esophageal and ear canal void spaces within the head. In particular, the void spaces alter photon transport in a manner not accounted for by a diffusion model. Ongoing research into more complex physical models and improvements to data collection methodologies will both help to alleviate these shortcomings.

The method developed herein describes a co-registration scheme that can use X-ray CT images acquired at a different time-point compared to the fluorescence measurements. Using advanced processing schemes, this concept could also be adapted to using mouse atlases instead of actual scans. However, the most obvious, and perhaps important, extension to this work will be the development of a single physical system for the collection of both FMT and CT data sets. As presented here, registration between the two is avail-

able only because the skull structure of each mouse presents a relatively straightforward rigid transformation, where the internal spatial relationships are maintained regardless of animal orientation or soft tissue distortion. Data collected concurrently or under identical placement could eliminate image these registration issues, enabling more accurate performance and the potential extension of the methods to areas other than the animal head, for example the chest cavity and abdomen, enabling the use of multi-modal FMT for preclinical imaging of virtually any fluorescent target. Even in its present form however, the method developed could complement stand-alone free-space FMT systems and offer a significantly higher degree of performance.

Another potential avenue for improvement of this technique is to automate the fragmentation of the original segments. Here, it was done geometrically to obtain segments which yielded a numerically stable reduced problem. If the heuristic approach presented here was instead replaced by an explicit optimization problem, issues of consistency and repeatability could potentially be reduced. With appropriately parameterized shapes and segments, techniques such as clustering algorithms could potentially aid in this process. The overall goal of such a procedure would be to create segments whose effects upon the diffuse forward model are as different from one another as possible, while maintaining reasonable shapes and minimizing the number of subsegments.

In conclusion, multimodal fluorescence imaging of tissue offers the potential for significant improvements in image fidelity and *in vivo* quantification over conventional fluorescence tomography methods. With the advent of new probe technologies and transgenic platforms utilizing fluorescence proteins, one could envision multi-modal FMT used to perform rapid and high-throughput visualization of neurodegenerative biomarkers. Optical sampling of human brain parameters has also been showcased in several studies [125,126], further pointing to the possibility for clinical propagation of the technology presented here. While it is not foreseen that the entire human brain can be visualized with optical methods in clinical settings, it would be nevertheless feasible to probe cor-

tex biomarkers at depths up to 3-5 cm, which could be sufficient for obtaining diagnostic information or clinical information on therapeutic efficacy.

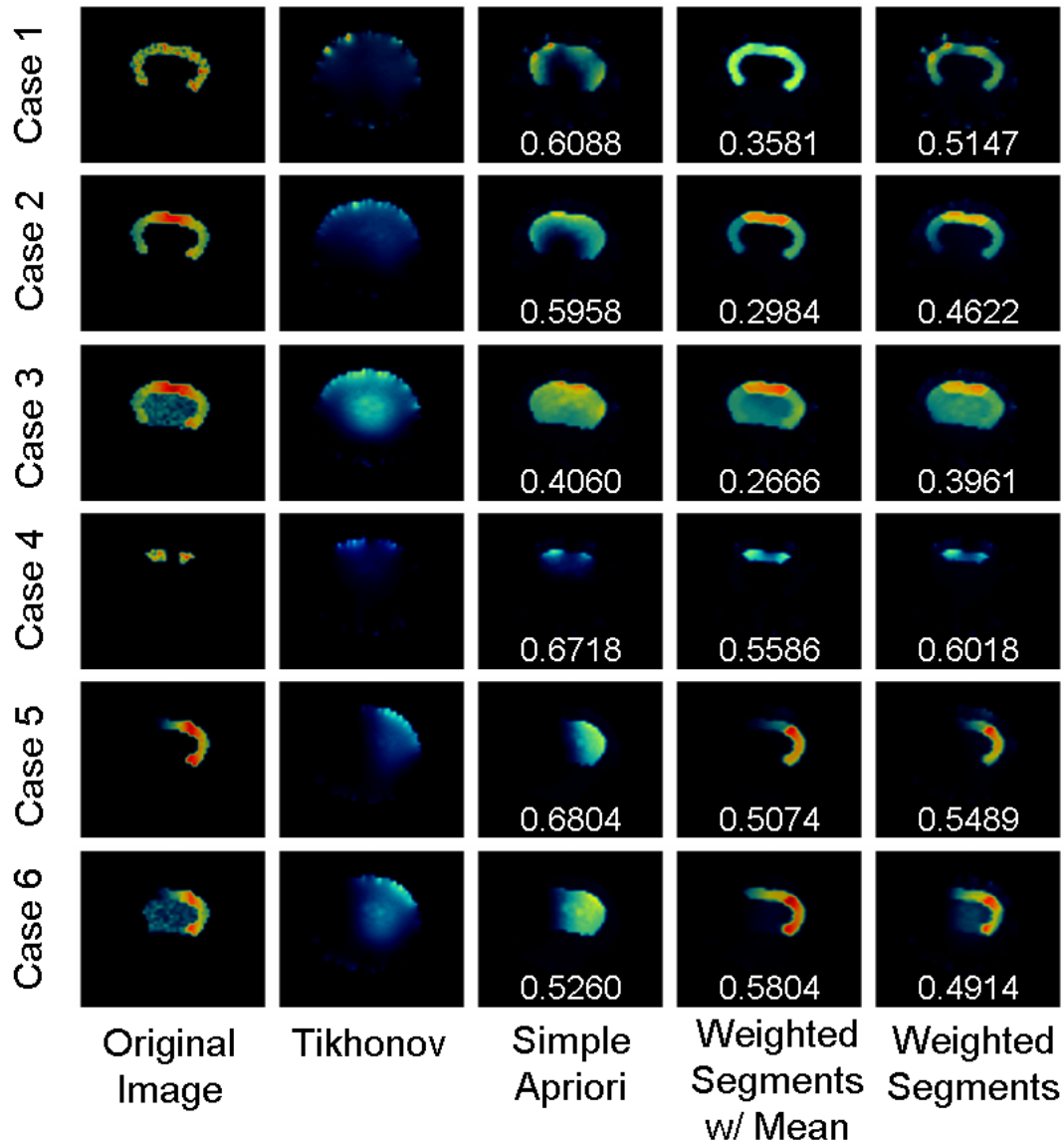


Figure 5.6: **Simulated Imaging Reconstructions** Ground truth is shown in the first column, while results using Tikhonov regularization with the identity matrix are in the second column. The simple *a priori* technique in the third column restricts the solution to either brain segment. The fourth and fifth columns show reconstructions using the techniques developed in Section 3, with and without explicitly imposing a mean value in the image prior model, respectively. Overlaid numerics indicate relative 2-norm error with respect to ground truth.

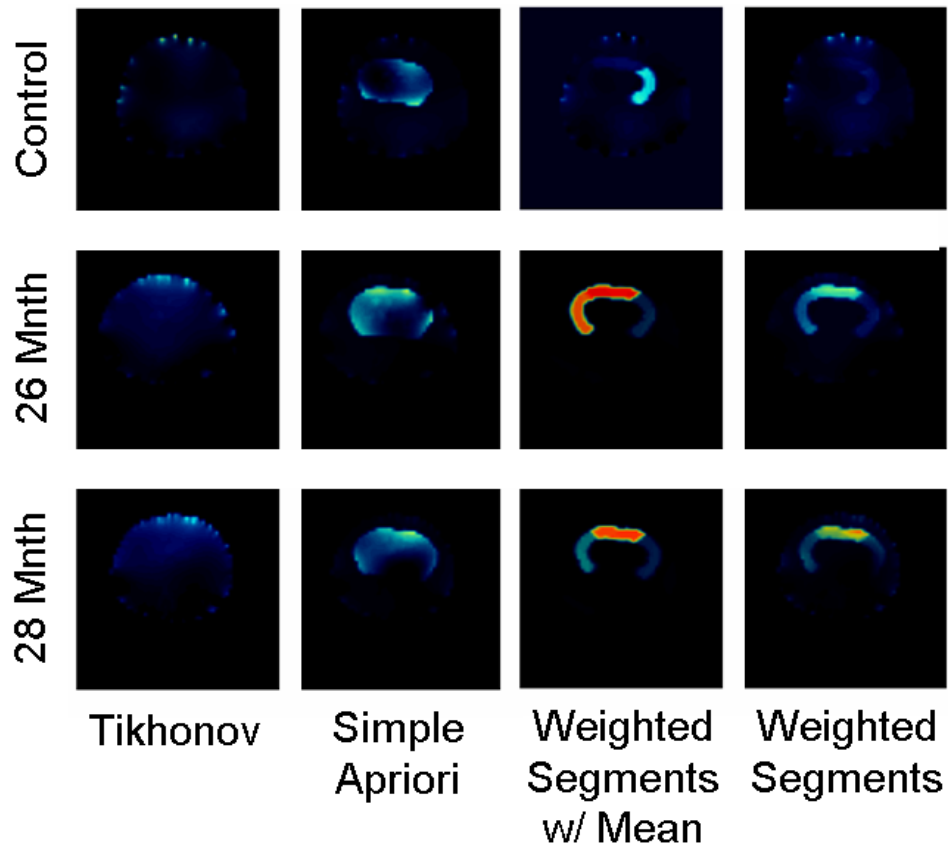


Figure 5.7: **In Vivo Imaging Reconstructions** A control animal is shown in the first row, while APP23 transgenic animals exhibiting amyloid- β plaques in the cortical segment (ages 26 and 28 months) are shown in the second and third rows. The first column shows reconstructions using Tikhonov regularization with the identity matrix, while the second columns using a simple *a priori* technique which favors solutions in the brain region. The third and forth columns show reconstructions using the techniques developed in Section 3, with and without explicitly imposing a mean value in the image prior model, respectively.

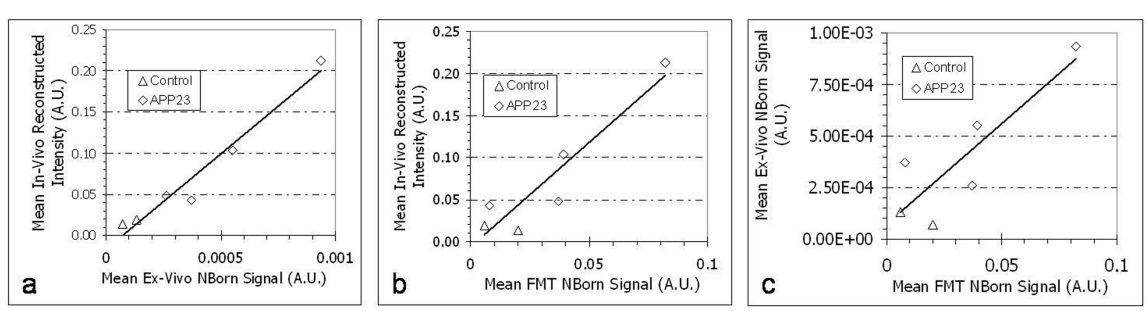


Figure 5.8: **In-Vivo vs Ex-Vivo Correlation** a) Correlation between *in vivo* reconstructed activity and normalized fluorescence intensity measured *ex vivo* (x-axis) ($R^2=0.9587$) . b) Comparison of mean input FMT data for a single projection (x-axis) with resulting FMT reconstructed intensity ($R^2=0.7422$) . The single projection was chosen with the camera positioned directly above the dorsal side of the head. c) Mean input FMT signal (x-axis) as in (b), correlated with measured ex vivo fluorescence activity ($R^2=0.8712$) .

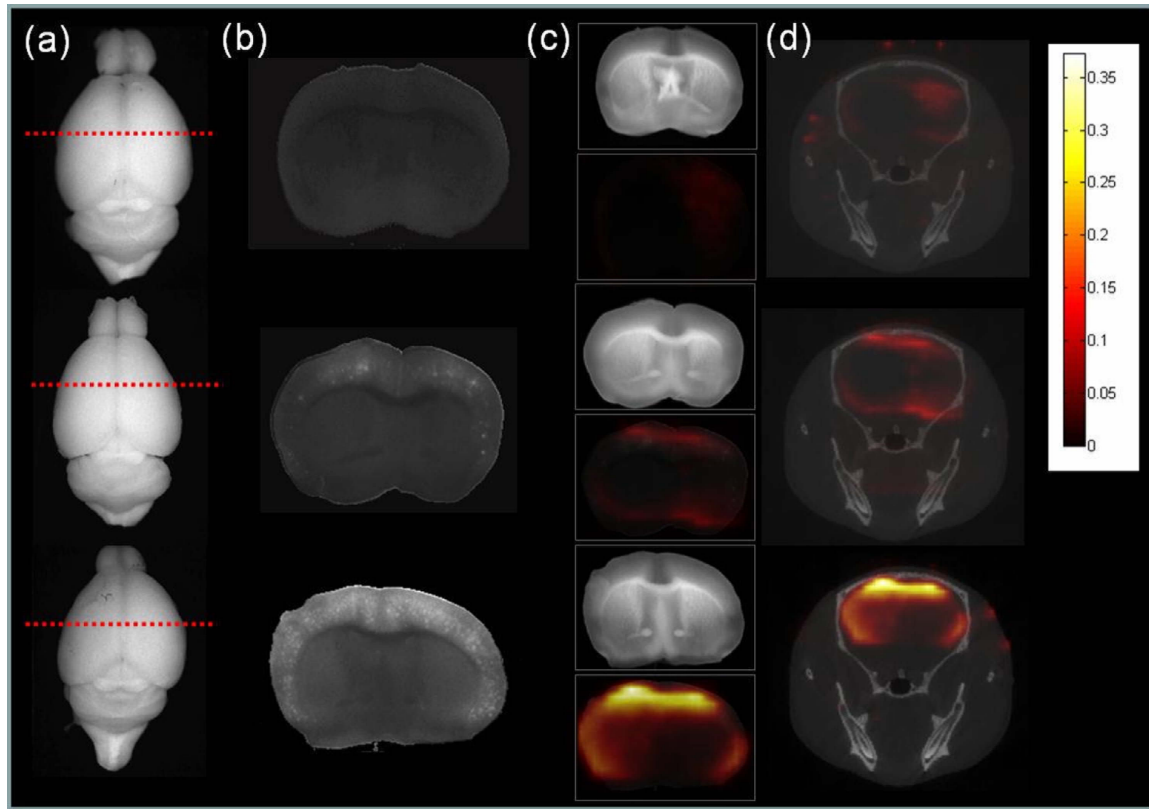


Figure 5.9: **In Vivo vs. Ex Vivo Imaging** Ex vivo versus *in vivo* FMT imaging comparison for 13 month old C57B/6 control mouse (First row) a 17 month old APP23 tg mouse (Second row) and a 26 month old APP23 tg mouse (Third row). (a) The first column shows full brain images in the excitation channel using a planar reflectance imaging system. The red dotted line denotes the approximate location corresponding to the slice shown in subsequent columns. (b) Planar reflectance images of normalized fluorescence from a single slice are then presented in the second column. (c) The third column presents planar images at the excitation wavelength (top) and FMT reconstructions overlaid on normalized planar fluorescence images (bottom). (d) *in vivo* multi-modal FMT reconstructions are shown in the final column for a slice corresponding to the same location as the *ex vivo* images, overlaid on a representative CT slice. All FMT reconstructions are scaled to the same colorbar.

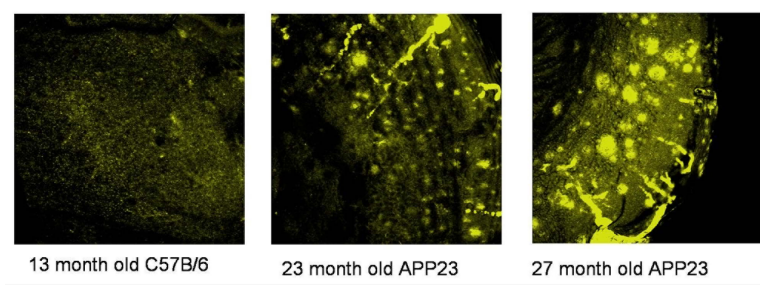


Figure 5.10: **Confocal Imaging** Representative confocal images of thick brain slices from a 13 month old C57B/6 mouse and two APP23 transgenic mice, at ages 23 and 27 months. In each image, amyloid- β plaques appear as bright areas.

Chapter 6

Robust use of Structural Information

6.1 Introduction

A final area of work examined for this thesis involves the development of techniques for using *a priori* structural information which are robust in the face of differences between the *a priori* anatomic segmentation used and the true fluorescence distribution which we are attempting to reconstruct. Such differences could arise from incorrect segmentation of the structural image or from fluorescence distributions which do not precisely conform to anatomic boundaries. As detailed below, the second case may arise because fluorescent probes react on a molecular rather than anatomic scale. The development of algorithms which are robust to such errors would be useful in improving algorithms for multi-modal imaging systems.

Considering the first situation more carefully, there are several reasons why the *a priori* boundaries in the structural image may be incorrectly located. For example, the structural and functional data sets may not be collected at the same time, or on the same apparatus. Registration is necessary in such situations to relate the data from each of the systems. If soft tissue deformation or other movement has taken place between the imaging sessions, perfect registration between the modalities will be difficult if not impossible. Even in situations where little or no deformation has taken place, the segmentation process itself can introduce errors. When accurate automated segmentation techniques are unavailable, manual segmentation must be done by experts, and the re-

sulting segmentation will be operator dependent. Additionally, if reconstruction methods can work well with sub-optimal segmentations, faster segmentation approaches, including manual segmentation by non-experts, could be used to simplify the process. Finally, such a technique could potentially allow the use of predetermined anatomical atlases in place of exact segmentations, enabling approximate structural information to be applied even in situations where it is not explicitly available.

As for the lack of precise correspondence between anatomy and function, in the case of optical fluorescence imaging, differences between the physical structure of the material and the structure of the fluorescence distribution may occur. Because FMT images molecular level activity, the parameters to be reconstructed are not necessarily bound by anatomic boundaries. The most interesting aspects of a functional image may be molecular activity which extends into the milieu beyond an organ boundary, or areas of an organ where the targeted activity is lower than the surrounding regions. As a specific example, the spread of cancerous growths may be preceded by biochemical changes in the surrounding tissue which have not yet manifested themselves as structural changes. The ability to accurately identify these regions *in vivo* would provide a valuable tool for researchers in the biological laboratory.

Below, we present two different solution approaches to this problem. The first uses differential equations as the basis for a Gaussian prior model of the image. While similar regularization schemes have been proposed on a tissue by tissue basis [30, 85], these existing methods do not permit information to be shared across boundaries. The inclusion of boundary conditions in the construction of a stochastically driven differential equation model, however, allows such interchange. This, in turn, allows fluorescence distributions which cross anatomic boundaries to be more accurately reconstructed.

The second method uses the sparsity inducing properties of 1-norm minimization as applied to underdetermined inverse problems. The *a priori* boundary locations are used to construct a series of vectors normal to the boundary, along which the 1-norm

of the gradient is computed. By extending these vectors to some limited extent on either side of the *a priori* boundary, the boundary's final location is constrained within a defined uncertainty region. To prevent artifacts from arising outside of the boundary region, least squares smoothing using a Laplacian operator is applied within the “trusted” regions away from this uncertain boundary. We apply this technique to a limited angle Radon type problem in place of FMT, because the underdetermined nature of the problem allows for an infinite number of valid solutions, and thus explicit 1-norm minimization, as discussed below.

6.2 Differential Equation Based Covariance Models

As stated in §2.1, a well known manner in which the regularized solution:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{W}\mathbf{x} - \mathbf{u}\|_2^2 + \lambda^2 \|\mathbf{L}(\mathbf{x} - \boldsymbol{\mu}_x)\|_2^2 \quad (6.1)$$

can be interpreted is as a maximum a posteriori (MAP) solution given a Gaussian image prior for \mathbf{x} with mean vector $\boldsymbol{\mu}$ and covariance matrix $(\lambda^2 \mathbf{L}^T \mathbf{L})^\dagger$, where \dagger indicates the pseudoinverse [72]. Typically however, the construction of \mathbf{L} is not interpreted in this manner, and also $\boldsymbol{\mu}$ is frequently assumed to be a vector of zeroes. Construction of \mathbf{L} is carried out to explicitly minimize image traits such as the local gradient or Laplacian. While this can quickly encode simple image constraints, more complex relationships are possible for which the appropriate regularization structure may not be readily evident.

To enable more complex regularization for diffusion and other inverse problems, we propose here an alternative method for the generation and analysis of correlation (regularization) matrices. Rather than explicitly construct the regularization matrices, we allow them to arise implicitly from the construction of a differential equation governing the behavior of the solution image [72, 127]. Specifically, an anatomically based differential equation is written, using a noise source of known statistical characteristics as the driving term.

Parameters for the differential equation are obtained using the low dimensional inverse solution described in Chapter 5. To incorporate relationships between anatomical regions, boundary conditions are established for each pair of neighboring regions. Once this has been done, a discretized version of the equation can be written as a linear system, which is then inverted to obtain an expression for the image prior in terms of the discretized differential operator and the driving noise process. This type of approach has previously been applied to oceanographic monitoring applications, where sparsely sampled data points are used to estimate behavior across an entire region [128].

6.2.1 From Differential Equation to Covariance Matrix

The central concept of this approach is that a linear differential system driven by a stochastic noise process can be discretized and written as:

$$\mathbf{A}\mathbf{x} = \mathbf{w}. \quad (6.2)$$

Here, the matrix \mathbf{A} encodes the differential operator with boundary conditions, \mathbf{x} is the discretized image, and \mathbf{w} is the driving noise process. If \mathbf{w} is a Gaussian random vector with $\mathbf{w} \sim N(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$, then:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{w} \quad (6.3)$$

immediately implies a stochastic model for \mathbf{x} of:

$$\mathbf{x} \sim N(\mathbf{A}^{-1}\boldsymbol{\mu}_w, \mathbf{A}^{-1}\boldsymbol{\Sigma}_w\mathbf{A}^{-T}). \quad (6.4)$$

When used as a prior probabilistic model for \mathbf{x} in the tomographic inverse problem, a maximum a posteriori (MAP) estimation problem can then easily be constructed as a linear least squares minimization:

$$\hat{\mathbf{x}} = \arg \min_x \|\mathbf{W}\mathbf{x} - \mathbf{u}\|_2^2 + (\mathbf{x} - \mathbf{A}^{-1}\boldsymbol{\mu}_w)^T (\mathbf{A}^{-1}\boldsymbol{\Sigma}_w\mathbf{A}^{-T})^{-1} (\mathbf{x} - \mathbf{A}^{-1}\boldsymbol{\mu}_w) \quad (6.5)$$

which has the solution:

$$\hat{\mathbf{x}} = \left(\mathbf{W}^T\mathbf{W} + (\mathbf{A}^{-1}\boldsymbol{\Sigma}_w\mathbf{A}^{-T})^{-1} \right)^{-1} \mathbf{W}^T(\mathbf{u} - \mathbf{A}^{-1}\boldsymbol{\mu}_w). \quad (6.6)$$

Note that this is analogous to (2.4) if the matrix \mathbf{L} is constructed such that $\mathbf{L}^T \mathbf{L} = (\mathbf{A}^{-1} \Sigma_w \mathbf{A}^{-T})^{-1}$.

In multi-modal FMT, the structure upon which this differential model is based is constructed at the higher resolution of the structural image rather than the lower final FMT resolution. Thus, if a differential model constructed at the structural resolution is denoted:

$$\mathbf{x}_{CT} \sim N(\mathbf{A}^{-1} \boldsymbol{\mu}_w, \mathbf{A}^{-1} \Sigma_w \mathbf{A}^{-T}) \quad (6.7)$$

The matrix \mathbf{C}_{CT}^{FMT} , defined in §5.2.1 (which transforms an image from the CT image scale to the FMT image scale), will be necessary to obtain a stochastic prior at the FMT resolution as:

$$\mathbf{x}_{FMT} \sim N(\mathbf{C}_{CT}^{FMT} \mathbf{A}^{-1} \boldsymbol{\mu}_w, \mathbf{C}_{CT}^{FMT} \mathbf{A}^{-1} \Sigma_w \mathbf{A}^{-T} (\mathbf{C}_{CT}^{FMT})^T) \quad (6.8)$$

In this manner, any image model derived from a PDE can be encoded at the resolution of the structural image, where a one to one correspondence exists between image voxels and anatomic regions, and subsequently shifted to the FMT image resolution. This is the general form of our approach. The details of the construction of the differential equation have significant bearing upon the traits of the resulting covariance matrix, and hence the regularization of the FMT inverse problem.

6.2.2 An Example Implementation

We now present an example of this approach, using a general image model with the potential for wide applicability. Within each anatomical region Ω_m , we assume that the solution $\mathbf{x}(r)$ should obey the differential equation:

$$\nabla^2 \mathbf{x}(r) - \gamma_m^2 \mathbf{x} = \mathbf{w}_m(r) \quad (6.9)$$

where $\mathbf{w}_m \sim N(0, \sigma_{m1}^2 \mathbf{I})$ is a zero mean Gaussian white noise process with variance σ_1^2 .

This differential equation was chosen as it encourages general smoothness within each

region, and thus has the potential for applicability to a wide range of imaging problems. At the internal boundaries between each region, we impose the constraint:

$$\beta_{p,q} \hat{n} \cdot \nabla f(r) = \beta_{p,q} \mathbf{v}_{pq}(r) \quad (6.10)$$

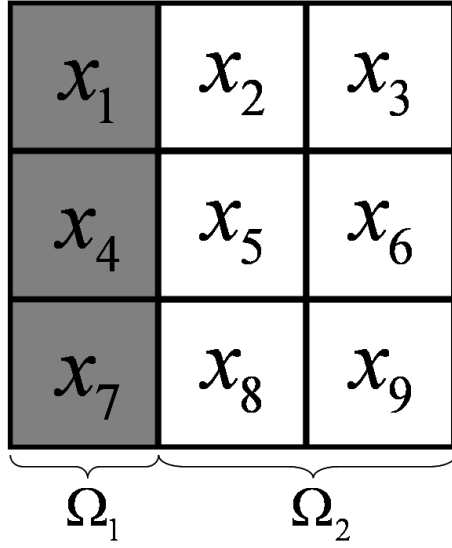
where $\mathbf{v}_{pq} \sim N(\boldsymbol{\mu}_2, \Sigma_{pq}^2)$ is a second vector of independent Gaussian variables, with mean vector $\boldsymbol{\mu}_2$ and nonuniform diagonal covariance Σ_{pq}^2 . **A zero boundary condition is used at the external boundary of the medium.** This choice of internal boundary condition will encourage a jump at the boundary between anatomic regions, with the size of the jump dependent upon the construction of the probability model of \mathbf{v}_{pq}

The parameter $\beta_{p,q}$ appears on both sides of the boundary condition because the left hand side of the boundary condition will be used in the construction of the differential operator, while the right will be incorporated into the driving noise process. Thus the value of $\beta_{p,q}$ scales the entire boundary condition so it is properly weighted against the primary differential operator. Mean values for $\mathbf{v}_{pq}(r)$ are computed as the difference between the mean values for the two adjoining regions, as determined using the low-dimensional parameterized solution. Variances for both $\mathbf{w}_m(r)$ and $\mathbf{v}_{pq}(r)$ are also computed from the low dimensional parameterized solution, using the approach detailed in the preceding chapter.

As illustrated in Figure 6.1, the above differential equation is discretized in 2D using an 4-neighborhood for both the primary differential equation and the boundary condition. Elements of the matrix representing the discretized equation are thus generated as:

$$\mathbf{A}_{i,j} = \begin{cases} -\gamma_p^2 - \sum_{k \neq i} \mathbf{A}_{i,k} & : & i = j \\ 1 & : & x_j \in \text{Neighbors}_4(x_i) \text{ \& } \{x_j, x_i\} \in \Omega_p \\ \beta_{p,q} & : & x_j \in \text{Neighbors}_4(x_i) \text{ \& } \{x_j\} \in \Omega_p, \{x_i\} \in \Omega_q \end{cases} \quad (6.11)$$

This allows an equation for the image \mathbf{x}_{CT} to be written in terms of the noise processes



$$(\gamma^2 - 3)x_5 + x_2 + x_6 + x_8 = 0$$

$$\beta(x_5 - x_4) = \beta(m_{\Omega_2} - m_{\Omega_1})$$

Figure 6.1: **Construction of Differential Regularizer** Shown here is a 3×3 voxel region about the voxel x_5 . Here, the voxels $\{x_1, x_4, x_7\}$ lie within region Ω_1 , while the remaining voxels lie within region Ω_2 . The mean values for each region are denoted m_{Ω_1} and m_{Ω_2} , respectively. To construct the row of the differential operator corresponding to x_5 , a 4-neighborhood is employed. The differential operator within Ω_2 results in the first equation written above, while the boundary condition results in the second. The final equation to be satisfied at x_5 is then the sum of these two equations.

\mathbf{w}_m and \mathbf{v}_{pq} :

$$\mathbf{x}_{CT} = \mathbf{A}^{-1}(\mathbf{w}_m + \mathbf{v}_{pq}) \quad (6.12)$$

where the construction of \mathbf{A} as a discretization of a well defined differential operator with boundary conditions ensures its invertibility. This, in conjunction with $\Sigma_w = \sigma_1^2 \mathbf{I} + \Sigma_2$ and (5.4), give us the model for the image at the FMT image scale, \mathbf{x}_{FMT} . This can then be used to write a maximum a posteriori inverse solution as:

$$\begin{aligned} \hat{\mathbf{x}} = \arg \min_x & \|\mathbf{u} - \mathbf{W}\mathbf{x}\|_2^2 + \\ & \lambda^2 (\mathbf{C}_{CT}^{FMT} \mathbf{A}^{-1} \boldsymbol{\mu}_w - \mathbf{x}_{FMT})^T (\mathbf{C}_{CT}^{FMT} \mathbf{A}^{-1} \Sigma_w \mathbf{A}^{-T} (\mathbf{C}_{CT}^{FMT})^T)^{-1} (\mathbf{C}_{CT}^{FMT} \mathbf{A}^{-1} \boldsymbol{\mu}_w - \mathbf{x}_{FMT}). \end{aligned} \quad (6.13)$$

6.2.3 Effects of Parameter Selection on Covariance

Looking at (6.9) and (6.10), it is clear that the values of γ^2 and β chosen for each region will have a direct effect upon the resulting structure of the correlation matrix. The parameter γ^2 will adjust the amount of smoothing applied to each region, while β will control correlation between the regions. In order to better understand the effects boundary conditions have upon the covariance structure of the image prior, as well as the role of the selection of β and γ^2 , we look now at a one dimensional version of the same differential equation, incorporating two physical regions, and the boundary condition as described above. Allowing β and γ to take on a range of values, we can more closely examine the effects of each.

Allowing β to take on a range of values, while holding fixed $\gamma = 0.5$, covariance matrices for the voxels along the one dimensional space can be constructed as:

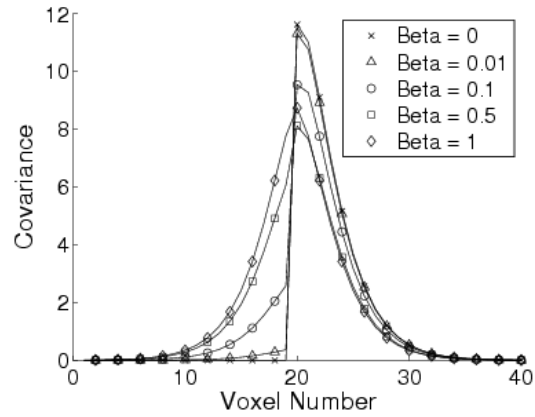
$$\Sigma = \mathbf{A}^{-1} \mathbf{A}^{-T} \quad (6.14)$$

where \mathbf{A} is the matrix encoding the one dimensional differential operator with boundary conditions. Figure 6.2(a) shows the 20th row of these matrices, for several values of β , plotting the covariance between voxel 20 and all other voxels for each β value.

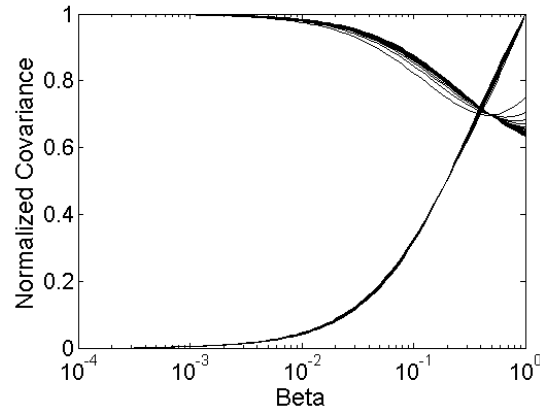
What can be seen in this graph is the effects of introducing correlation across the boundary between the two regions. When $\beta = 0$, there is no relationship between the two regions, and the covariance immediately drops to zero when crossing the boundary. As the value of β is increased, the covariance present between voxel 20 (lying in the right hand region), and voxels in the left hand region, begins to increase. For values of β between 0 and 1, there still exists a drop in correlation when crossing the boundary. Once $\beta = 1$, however, the boundary in a sense disappears. This occurs because the structure of the covariance matrix is dependent only on the left sides of the differential operator and boundary condition. When $\beta = 1$, the left side of the boundary condition will be identical to the differential operator that would be present in the absence of a boundary. The right side of the overall equation (differential operator plus boundary condition) will however be different as it represents the stochastic process driving the system. When the variances of the left-hand and right-hand regions are identical, the resulting covariance plot will be symmetric as seen in 6.2(a). If the variances are not identical, the structural symmetry of the plot will remain, but each side will be scaled according to the variance of its corresponding region.

This plot can also be used to examine how changes in β affect the structure of the resulting covariance matrix. For $\beta = 0$, the covariance matrix will take on a block diagonal structure, with one block for each of the two physical regions. As β is increased, non-zero covariances will arise in the off-diagonal blocks, evidenced by the non-zero covariances between voxel 20 (in the right hand region), and voxels 1-19 (in the left hand region) seen in Fig. 6.2(a).

Figure 6.2(b) shows a related plot of the changes in covariance as the value of β changes. Forty lines are plotted in the graph, one line for the covariance between voxel 20 and each of the 40 voxels. Each line is normalized such that its maximum value is equal to 1, yielding a measure of the relative change in covariance as β is varied. These forty lines can be seen to fall into two distinct groups, corresponding to whether the associated



(a)



(b)

Figure 6.2: **Change in Covariance Structure as a Function of β :** a) Plot of covariance of voxel 20 with all other voxels, for a range of β values, given $\gamma^2 = 0.5$. b) Plots of the relative change in covariance between voxel 20 (right hand region), and all other voxels, as a function of β . Plots are normalized such that their maximum value is equal to 1. Note that plots are clustered into two groups. The lower, exponentially increasing group, corresponds to the covariance between voxel 20 and voxels within the alternate (left-hand) region. The upper plot shows covariance between voxel 20 and all other voxels within the right-hand region. Note that most voxels have a relative decrease in covariance as β is increased. However several plots, corresponding to those voxels close to the boundary, show a decrease followed by an increase in covariance as β is increased.

voxel falls into the right-hand (voxels 1 through 19) or left-hand (voxels 20 through 30) region. Those voxels falling into the left-hand region have covariance plots which form the lower line of the graph and which, as β is increased, increase correspondingly in approximately the same relative manner (ie- if a particular change in β results in a 50% increase in covariance between voxel 20 and voxel 17, then the increase in covariance between voxel 20 and each of voxels 1-19 will be approximately 50%).

The other group of voxels are those that fall within the same region (the right-hand region) as voxel 20. The covariance plots for these voxels are clustered together in the top line of the graph, and show covariances which roll off and decrease as β increases beyond approximately $1e^{-2}$. Interestingly, for those voxels clustered closely to one another, in the same region, and near the boundary, the decrease in covariance actually reverses itself as β approaches 1. The precise reason for this reversal is currently unknown, and will require further investigation.

Figure 6.3 shows the changes in covariance between voxel 20 and all voxels as a function of γ^2 , with β fixed equal to 1. What can be seen in this plot is that as γ^2 increases, the covariance becomes a function which decreases exponentially as a function of distance from voxel 20. This result compares well with previously published results for the case of a one-dimensional diffusion equation in an unbounded continuous space [128]. For that case, the covariance between any two points is a function only of the distance between those points, and can be expressed as [85, 128]:

$$C(r) = (1 + \gamma r)e^{-\gamma r}. \quad (6.15)$$

The suggested interpretation of this is that γ^2 can be seen as setting a “correlation distance”, which determines at which distance two points will still be significantly correlated.

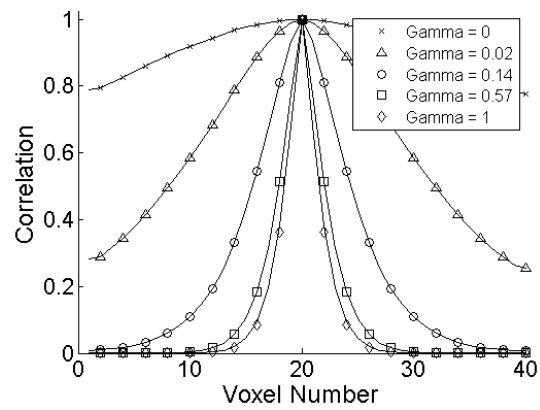


Figure 6.3: **Change in Covariance Structure as a Function of γ^2 :** Covariance between Voxel 20 and all others for a range of γ^2 values. Note that as γ^2 increases, the covariance plot becomes a double sided exponential,

6.2.4 Results

We now apply this regularization technique to two 2D examples simulating fluorescence within the murine chest cavity. This type of imaging target could arise, for example, when localizing inflammation within the lungs. To generate data for these examples, a finite element mesh was generated on a geometry extracted from an X-Ray CT of a mouse. Each anatomic region was assumed to be optically homogeneous, and individual optical properties were assumed known. The finite element solutions were used as Green's functions to compute simulated data sets using the normalized Born approximation. The model used for the inverse solutions was also computed from the finite element model, by downsampling onto a regular rectangular grid. Shot noise equal to 10% of the signal intensity was also added to the simulated data.

Figure 6.4 simulates a situation where the fluorescent inclusion is present primarily within the lung, but with regions of high intensity that extend beyond the *a priori* boundary of the lung. For this case, we have taken $\gamma^2(\Omega_j) = 1 \quad \forall \Omega_j$ and $\beta(\Omega_p, \Omega_q) = 0.25 \quad \forall \Omega_p, \Omega_q$. This corresponds to approximately a 40% reduction in correlation between regions as compared to the boundary-less case. In all images presented, the *a priori* boundary of the right lung is indicated by a solid red line. Reconstruction using Tikhonov regularization with the identity matrix results in Fig 6.4(b). While the general location of the fluorescent inclusions are correct, the image is quite smooth and doesn't accurately represent the underlying fluorescence distribution. Using a regularization scheme which applies a discrete approximation to a Laplacian operator to each region independently yields Fig 6.4(c). This image is clearly an improvement over the identity regularized result; however the reconstruction is still quite smooth, and lacks some of the structural detail of the original. Applying our differential equation based regularization scheme results in Fig 6.4(d). Here, the structure of the lungs is more accurately reconstructed than with the other approaches. Additionally, because of the incorporation of boundary conditions in the construction of the differential equation,

our technique is able to identify the fact that fluorescence extends beyond the boundary of the right lung.

Figure 6.5 shows the results of this method applied to another lung inflammation target, with significantly more complex structure. The results obtained with the Tikhonov and Laplacian regularization schemes are similar to those seen for the previous lung inflammation simulation, despite the significant differences between the two cases. The two reconstructions are heavily smoothed, and while the reconstructions both appear vaguely lung-like, they do not incorporate the detail seen in the original. Reconstructing using the differential regularizer with boundary conditions, however, results in the image seen in Fig. 6.5(d). Several improvements are notable in this image. First, there is a fluorescent inclusion reconstructed outside of the lungs, that is also indicated in the original image, but missing from the Tikhonov and Laplacian solutions. Second, the structure of the fluorescence distribution in the left lung is more accurately reconstructed. Finally, in the right lung, the region of low fluorescence at the right edge of the lung is better reconstructed. Each of these is a substantial improvement over the Tikhonov and Laplacian solutions.

6.3 1-Norm Based Floating Boundaries

While the approach presented in the preceding section allows for correlation between individual anatomic regions to be incorporated into the regularization scheme, the results still tend to be heavily smoothed by the differential operator, even at the boundaries. In some cases where *a priori* structural information is available, it may be known or assumed that a sharp boundary should be present within the functional image, at a location near the boundary in the anatomic image. To achieve this, we desire to have a single boundary location that is able to shift within some predetermined limits.

To develop such an approach, we examined a regularization scheme based on a combination of standard linear smoothing and use of the 1-norm of the image gradient

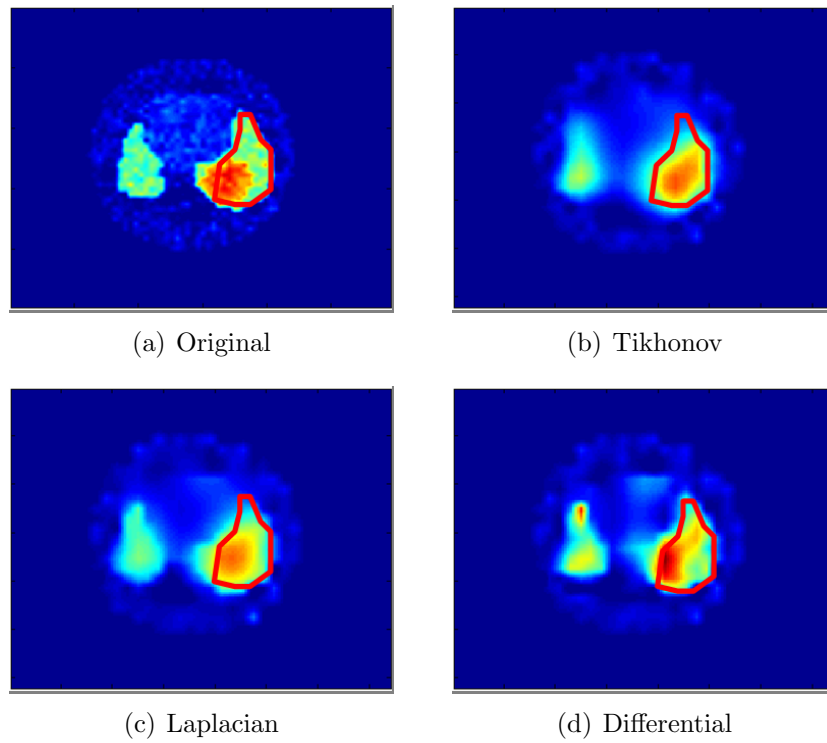


Figure 6.4: **Reconstructions of Simulated Fluorescence within an Artificial 2D Geometry:** a) Fluorescence Distribution used for Data Generation. Red line denotes boundary of lung region b) Reconstruction using Tikhonov Regularization with the Identity Matrix. c) Reconstruction using Laplacian Regularizer applied to each region independently. d) Reconstruction using Differential Equation Based Regularization Technique.

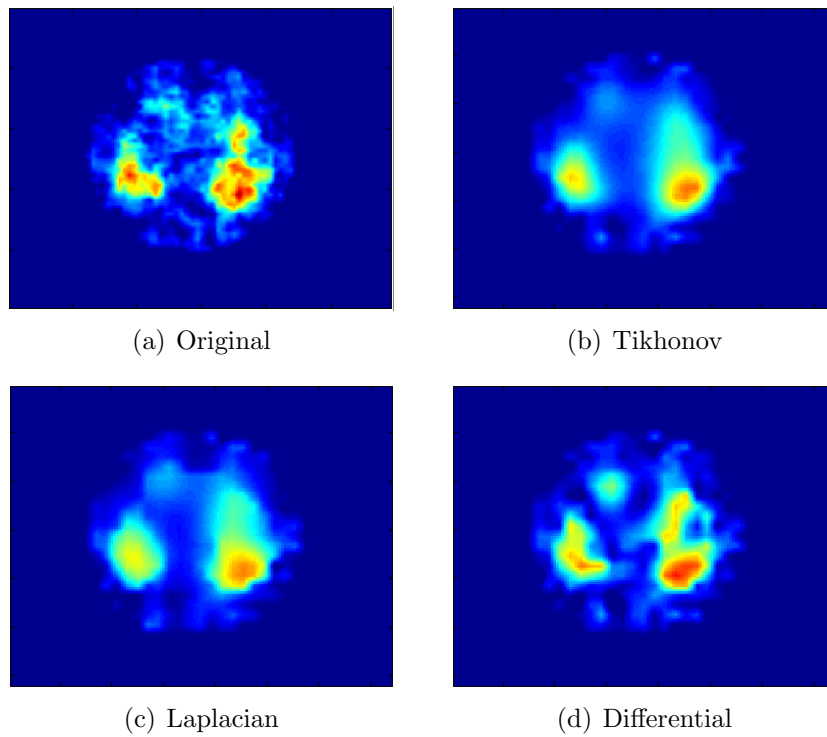


Figure 6.5: **Reconstructions using Differential Regularizer on a Complex Target:** a) Fluorescence Distribution used for Data Generation. b) Reconstruction using Tikhonov Regularization with the Identity Matrix. c) Reconstruction using Laplacian Regularizer applied to each region independently. d) Reconstruction using Differential Equation Based Regularization Technique.

in regions surrounding the *a priori* boundary locations. The 1-norm is known to encourage sparsity, and the conceptual idea is that applying the 1-norm to the image gradient in a region near the boundary will result in a single jump somewhere within that region. Specifically, we divide the reconstruction space into trusted and boundary regions. Trusted regions are those areas where the label assigned by the *a priori* segmentation can be trusted as correct. All other areas are defined to be boundary regions, where we examine the 1-norm of the gradient, taken along lines that are normal to the *a priori* boundary location.

As shown in Figure 6.6, the trusted and boundary regions were defined by introducing an uncertainty measure Δ , which defines the distance normal to the anatomical boundary at which the boundary of the trusted region lies. Within the trusted regions, regularization is applied to the 2-norm of the gradient. This is implemented as a matrix \mathbf{L} , and yields the penalty term $\|\mathbf{L}\mathbf{x}\|^2$. By regularizing the gradient of the image, smoothness will be encouraged within each of the trusted regions.

Voxels which do not fall within a trusted region are considered to be boundary voxels. The expectation is that if a straight line is traced from one trusted region through a boundary region to another trusted region, somewhere along that line a jump should be encountered in the image. The location of this jump will correspond to the boundary of the functional region, which may or may not be in the same location as the *a priori* anatomic boundary. Choosing the straight lines to be vectors normal to the *a priori* boundary, the uncertainty measure Δ is used to define how far in either direction one must traverse the line in order to guarantee to have passed from one region to another. Figure 6.6 diagrams how the trusted and boundary regions were computed, along with the construction of the boundary normal lines.

For each of these boundary normal lines, the 1-norm of the image derivative along that line was then computed, and the sum across all boundary normals taken as an image

penalty constraint, as:

$$\sum_{i=1}^{N_{normals}} \int_{\Omega} |\nabla \mathbf{x}| \delta_i(\mathbf{r}) d\mathbf{r}. \quad (6.16)$$

Here $\nabla \mathbf{x}$ is the gradient of the image, while $\delta_i(x, y)$ is a delta function supported on the line segment \mathbf{n}_i , the i^{th} vector normal to the *a priori* boundary location, extending a length Δ in either direction from that boundary. The integral is taken across the entire solution space Ω , with respect to the coordinate location \mathbf{r} . This penalty constraint can be implemented on a discrete voxelized space as:

$$\sum_{i=1}^{N_{normals}} \|\mathbf{D}_i \mathbf{x}\|_1 \quad (6.17)$$

where the matrix \mathbf{D}_i computes the appropriate first order differences along the normal vector \mathbf{n}_i between voxels in the pixelated image, while the 1-norm takes the absolute value and sums them together.

Similar to other 1-norm minimization work done for compressed sensing [129, 130], we look to solve this problem using convex optimization techniques. While general nonlinear methods such as steepest descent, Newton's method, or nonlinear conjugate gradient, could be applied, they do not leverage the convexity of the problem in \mathbf{x} . Because of this, we use the Matlab package CVX [131, 132] to solve the minimization:

$$\min \sum_{i=1}^{N_{normals}} \|\mathbf{D}_i \mathbf{x}\|_1 \quad (6.18)$$

subject to:

$$\|\mathbf{W}\mathbf{x} - \mathbf{u}\|_2^2 \leq \delta \quad (6.19)$$

$$\|\mathbf{L}\mathbf{x}\|_2^2 \leq \lambda^2. \quad (6.20)$$

The first constraint is a data fitting term, while the second applies a linear smoothness constraint on each trusted region. The parameter δ is taken to be the expected noise power in the measured signal, while λ^2 will act as a regularization parameter controlling the degree of smoothing within each of the trusted regions. For this proof of concept, we assume that the noise power is known.

6.3.1 Limited Angle Radon

One issue that arises with 1-norm optimization as a sparsity inducing measure is that the system being solved must be underdetermined. In applications such as compressive sensing [129, 130], the 1-norm is used in place of stronger sparsity-inducing norms such as p-norms for $0 < p < 1$. The ultimate sparsity inducing norm is the 0-norm, which unfortunately changes a convex optimization problem into a combinatorial optimization problem, thereby exponentially increasing the computational complexity. If, given a matrix \mathbf{W} describing the system, collected data \mathbf{u} , and an image \mathbf{x} to be recovered, a linear constraint of the form:

$$\mathbf{W}\mathbf{x} = \mathbf{u} \tag{6.21}$$

can be written (assuming for the moment that \mathbf{W} is $n \times n$), then use of the 1-norm becomes equivalent to use of the 0-norm [?]. Figure 6.7(a) pictorially describes this effect in 2 dimensions. Essentially, the surface defined by all solutions sharing the same 1-norm takes the shape of an n-dimensional diamond. The constraint $\mathbf{W}\mathbf{x} = \mathbf{u}$ defines a hyperplane in n-dimensional space upon which the solution must lie. As the n-dimensional diamond expands (corresponding to increasing the 1-norm of the solution), with high probability it will eventually be one of the tips of that diamond which first reaches the hyperplane defined by the linear constraint.

By contrast, when 1-norm minimization is applied to an overdetermined system, the set of valid solutions is no longer a hyperplane. By using a quadratic cost function for an overdetermined system, each point within the solution space will instead have a particular cost associated with it, with the isocontours of this cost function frequently being ellipsoidal in shape, as show in Fig 6.7(b). While the pointedness of the 1-norm ball is beneficial when fitting the solution to a linear hyperplane, all points on the surface of that 1-norm ball still share the same 1-norm value. When using the quadratic constraint, the cost function optimization will seek out the point on that surface where the quadratic function is minimized, and this point is highly unlikely to lie at one of the corners or

edges which define sparse solutions.

Given the low resolution nature of diffusive imaging, FMT problems are typically highly overdetermined. While they can be downsampled to yield an underdetermined problem, this typically involves throwing away large amounts of data. As an initial proof of concept example, we instead look at a limited view Radon type problem. This type of problem is encountered in situations such as breast tomosynthesis, where X-ray data is available for only a small number of angles over a limited angular view [133]. The result of this restriction is a significant increase in imaging artifacts as compared to full view CT with a high angular sampling density. By incorporating structural prior information into the inverse problem, we seek to reduce the number of artifacts present in the reconstructed image.

6.3.2 Results

Figure 6.8 results shows the results of applying this technique to a boundary location problem on a simple geometry. The original image used for data generation is seen in Fig. 6.8(a). This differs from the *a priori* assumed boundary of the region, which is circular, and visible as the border of the red region in the Laplacian regularized solution 6.8(c). The reconstruction in Fig. 6.8(b) was obtained using Tikhonov regularization with the identity matrix. For this example, the Radon problem was limited to a 60° field of view, with 6 evenly spaced angular views, and 75 samples per view. In this image, reconstruction artifacts resulting from the limited view nature of the problem are clearly visible. Introducing *a priori* knowledge to the problem, Fig. 6.8(c) was obtained by applying a Laplacian within each of the regions on either side of the *a priori* boundary location. Here, the reconstruction clearly shows the *a priori* located boundary, and completely fails to resolve any of the differences between that boundary and the true boundary of the object. Finally, Figs. 6.8(d)-6.8(f) present reconstructions obtained using our 1-norm based approach for varying values of λ . In each case, a hard boundary

is achieved, which more accurately represents the true location than was obtained with either previous approach. As the value of λ decreases, smoothness in the reconstruction correspondingly decreases, until at $\lambda = 0$, a piecewise constant solution is obtained. While an improvement on other approaches, the boundary location is still not correct in all locations. In particular, those portions of the boundary falling along the axis of the limited views are poorly reconstructed.

6.4 Conclusions

Multimodal techniques offer the potential for significant improvements in diffusion based optical imaging. Using information about anatomical structure, these methods can guide the reconstruction of fluorescence distributions to improve both spatial fidelity and quantitative accuracy. However when imaging fluorescence distributions, differences may exist between the *a priori* anatomical structure, and the structure of the underlying fluorescence distribution. We have presented here two new techniques which help to compensate for these errors by allowing the boundary location to shift from its original *a priori* location.

Our first approach uses differential equations and boundary conditions to construct regularization matrices which allow information to be exchanged across anatomic boundaries. Using a very general differential equation and boundary, we were able to obtain significant improvements in image quality over approaches which do not employ information transfer between anatomic regions. Given the ubiquitous nature of differential equations as models for physical phenomena, this approach has the potential to generate regularization schemes which are highly specific and particularly suited to each imaging target. If the target fluorochrome is known to behave in a specific pharmacological manner, it would be possible to write equations governing this and use them to regularize the solution. Thus, rather than penalizing based on some generalized characteristics of the solution image, the fluorescence tomography problem could be coupled to the underlying

biochemical processes. This potential for fusion between bodies of knowledge is what may make this differential equation approach to regularization a useful tool for multimodal biomedical imaging.

As a potential method for encouraging a jump at the boundary, our second method employed convex optimization and the use of the 1-norm as a regularization constraint. Rather than minimize the 1-norm of the gradient across the entire image, our approach used the *a priori* boundary locations to construct boundary normals along which the 1-norm was computed. This constrained the boundaries to lie within some predetermined distance of their *a priori* location. While the results obtained are still very preliminary, they suggest that future work may be able to employ such an approach to more accurately localize the boundary in the presence of errors in the prior knowledge.

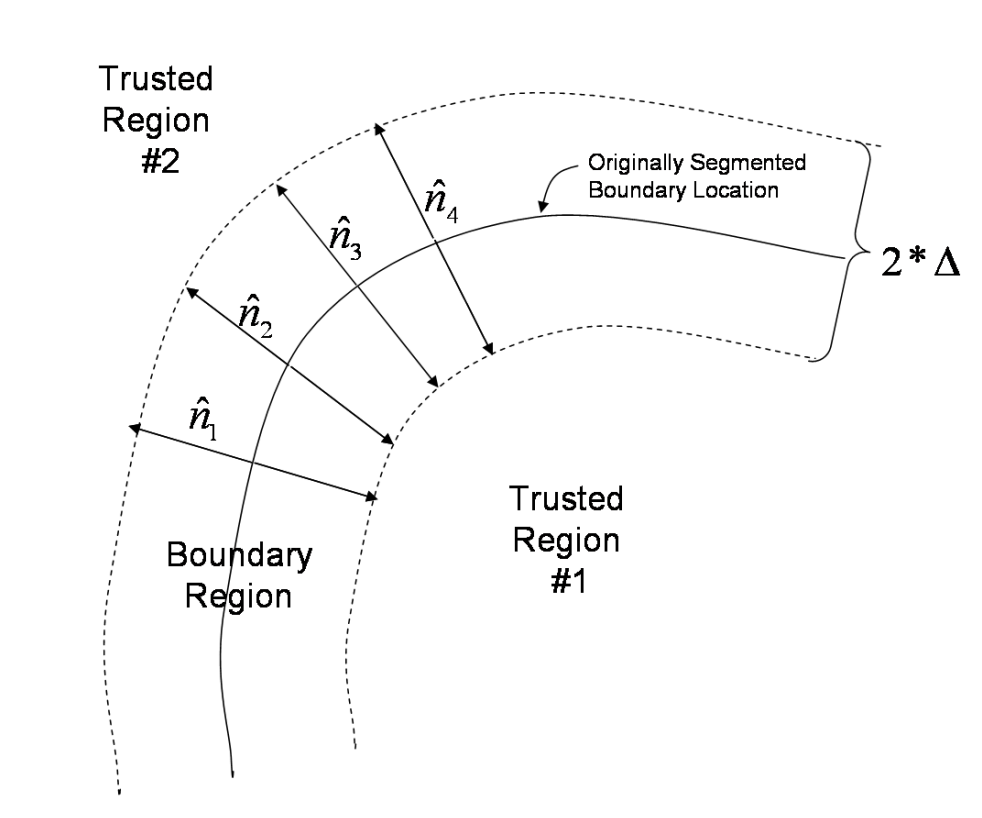


Figure 6.6: **Diagram of 1-norm Boundary Regularizer** At each of a number of points along the boundary denoted by the heavy black line, the normal vector was computed. Using this direction, and stepping some distance Δ in either direction from the boundary, the integral of the 1-norm of the derivative of the image along that line was computed. The summation of all such 1-norms was then used as a penalty term in the inversion process.

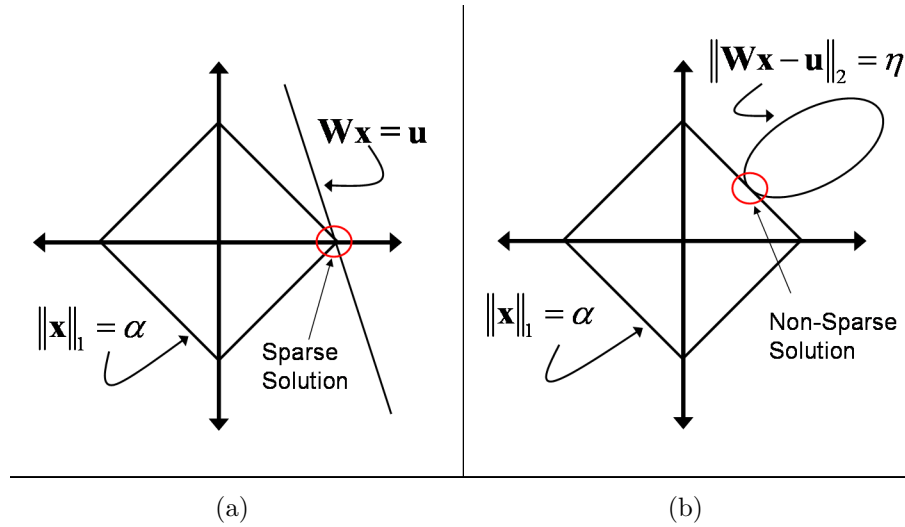


Figure 6.7: **The 1-norm and Sparse Optimization** a) By minimizing the 1-norm when solving a system subject to some underdetermined linear constraint $\mathbf{W}\mathbf{x} = \mathbf{u}$, the minimum norm solution which intersects the constraint plane will be a sparse solution, with high probability. b) If the linear constraint is overdetermined, an approach minimizing the norm $\|\mathbf{W}\mathbf{x} - \mathbf{u}\|$ needs to be taken, which likely results in a non-sparse solution.

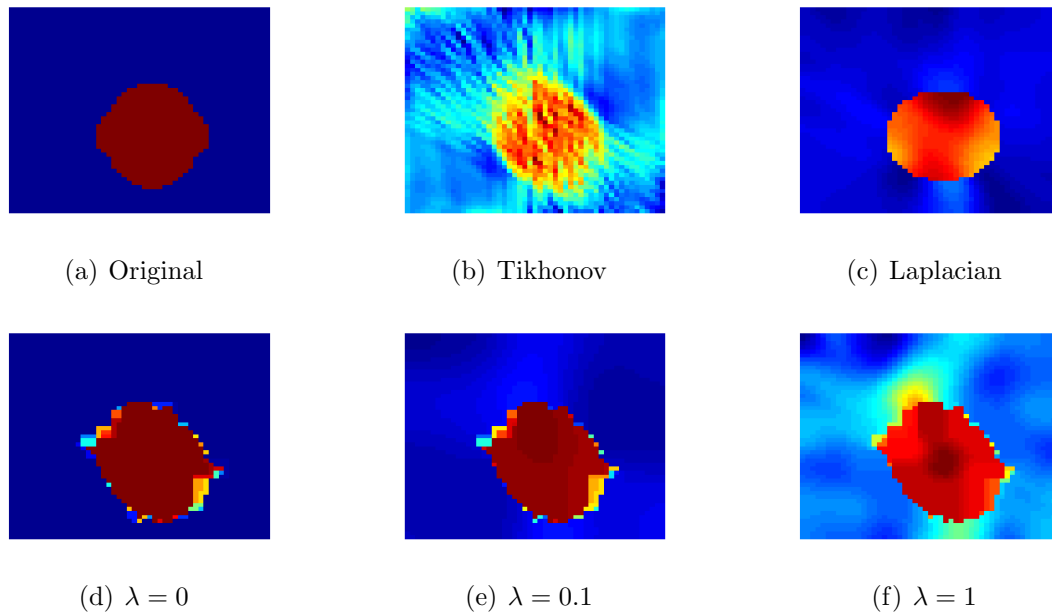


Figure 6.8: **Results Using Boundary Normal 1-Norm Regularization:** a) Fluorescence Distribution used for Data Generation. b) Reconstruction using Tikhonov Regularization with the Identity Matrix. c) Reconstruction using Laplacian Regularization on each region independently. d-f) Reconstructions using 1-norm regularization. d) $\lambda = 0$ e) $\lambda = 0.1$ f) $\lambda = 1$.

Chapter 7

Conclusions

In this thesis, we have presented approaches which use statistical signal processing techniques to improve reconstruction algorithms for fluorescence molecular tomography (FMT). In Chapter 1.2, we presented existing work on optical tomographic techniques, and detailed the imaging systems which generated the data sets processed in this thesis. Research into system development is continuing to improve the quality of both single and multi-modal datasets, which will in future contribute to further improved imaging capabilities. Chapter 2 presented an introduction to inverse problems, regularization, and applicable algorithmic methods. Future work into the effects of constrained versus non-constrained solution methods for FMT will help determine the best algorithmic approaches for solving these problems, and may be aided and accelerated by work such as the modified L-Curve presented in Sec .

In Chapter 3, we examined the normalized Born ratio from a statistical perspective, building up a probability model from experimental data, and resulting in several solution methods based on a sequence of approximations. These methods were shown to yield superior reconstruction results as compared to existing threshold based techniques. Furthermore, by eliminating the manual selection of thresholds, this technique allows for optimal use of all available data, and reduces variability among reconstructions from multiple datasets. This in turn allows for more accurate comparison of reconstruction images.

Two assumptions of independence were made in the course of our derivations. The

first was the independence of each individual fluorescence and excitation measurement from all other measurements of the same type. The second assumed that each fluorescence measurement was independent of its associated excitation measurement. Neither of these assumptions is strictly true, due to errors in the assumed physical model. Were the model perfectly accurate, the measurement noise at each detector would be independent of one another. However, differences between the model and the true physical process can be considered as additional noise in the system. Given the smoothness of diffusion photon fields, this noise will contain some degree of spatial correlation. If this model error could be quantified and accounted for, the noise covariance matrix would become non-diagonal, and better able to compensate for the shortcomings of the approximate linear model. As model error is highly dependent upon the physical system being observed, approximation will likely be necessary in order to develop techniques which can be applied to multiple imaging targets. Additionally, the utility of such models would need to be evaluated, to determine whether the required additional complexity would result in sufficient gains in accuracy to justify their use.

Chapter 5 used a single core concept for multi-modal signal processing to define two new regularization methods, as well as several potential avenues for future research. The central idea of this chapter is that when imaging with multi-modal data sets, parameterized image models can be constructed whose parameters can then be estimated directly from the collected data. In this way, regularization terms are defined which are customized to each individual data set. This allows for the creation of regularization terms that require fewer parameters to be defined *a priori*, and allow for differentiation between test and control animals without such prior knowledge being explicitly incorporated into the inversion scheme.

This technique was then applied to an *in vivo* imaging study which clearly showed the utility of multi-modal FMT-CT. Studying the development of amyloid- β plaques in the brains of transgenic mice provided an experimental study which could not have been

performed without multi-modal techniques. Incorporating correlative *ex vivo* imaging, this Alzheimer’s imaging study showed that FMT can achieve quantitative results *in vivo* in difficult imaging scenarios, extending the capabilities of FMT as a tool for advancing biological science.

Potential future work offers several interesting expansions upon this technique. While the current two step multi-modal approach uses a simple relationship between mean and variance, further developments could provide ways for estimating all parameters of the structured image model directly from the collected data. If the structure of the covariance matrix is known to within a small number of scaling coefficients (i.e. variances on each of the physical regions), an approach such as that diagrammed in §5.5 could be applied to explicitly estimate the variances, rather than using the heuristic approach presented in Chapter 5. By computing statistically robust variance estimates directly from the data, the resulting prior models will more accurately reflect the underlying fluorescence distribution, and likely result in improved reconstruction capabilities.

Another area in which the two step multi-modal technique could be improved is a more analytic method of choosing the subsegmentation of the original anatomic regions. In this thesis, the cortical region of the brain was divided into three subsegments to improve the numerical properties of the resulting reduced dimensional inverse problem. This division was done in a heuristic manner, and evaluated using the average angle between the columns of the resulting reduced matrix. Future work could examine this process more closely. An initial goal could be the development of an improved metric for judging the quality of each potential subsegmentation. By taking into account both the number of subsegments and the shape of each subsegment, as well as the resulting effects upon the reduced forward matrix, a minimization scheme could be developed which allows selection of an optimal subsegmentation.

Our two step approach as presented in Chapter 5 can in some sense be considered a two resolution multi-grid method. If instead a series of such subsegmentations were

developed, corresponding to a change in average subsegment size, existing research on multi-grid methods could be used to develop algorithms which allow further improved reconstruction results. Current multi-scale methods typically involved a sequentially decimated grid pattern [134] which is not constrained by anatomical boundaries. By incorporating multi-modal anatomical information into the construction of the solution spaces at each resolution level, improvements in both reconstruction speed and accuracy may be available that are unachievable with fixed grid approaches.

Significant additional gains will also be made through the development of improved imaging apparatus, and construction of dedicated multi-modal imaging systems. The multi-modal data used in this thesis was collected on multiple imaging systems, requiring the use of registration to combine the two datasets. By constructing a single imaging system capable of collecting both functional fluorescence data and structural information in parallel, the effects of misregistration and soft tissue deformation will be minimized. These reductions in error will likely result in significant improvements in the reconstructions seen with all multi-modal approaches applied to the data collected.

Finally, Chapter 6 offered two approaches to help provide robustness to differences between the anatomical structure acquired from multi-modal data, and the structure of the underlying fluorescence distribution. Using a differential equation with boundary conditions as the prior model for the fluorescence, the first approach constructed a model which allowed information to be passed between anatomical regions, thereby helping to correct for improperly located boundaries. The second approach used a 1-norm constraint along lines normal to the boundary to encourage a single large jump within some distance of the *a priori* boundary location.

Future work using differential equation based image prior models could allow inverse problems to easily incorporate knowledge from other areas of signal processing, physics, chemistry, and biology. Wide bodies of research have been done on using differential equations to model a wide range of phenomena. By working with biologists and

biochemists to better understand the physical processes being imaged, it may be possible to create prior models which more accurately model the physical and pharmacological behavior of fluorochromes, allowing for corresponding increases in the accuracy of *in vivo* FMT imaging.

As further advancements make fluorescence molecular tomography more accurate, reliable, and repeatable, the technology will further disseminate throughout the biological laboratory. Replacing existing planar fluorescence imaging systems, FMT will provide better quantitative representation of *in vivo* fluorescence activity in whole animals.

Bibliography

- [1] Fred S. Wouters, Peter J. Verveer, and Philippe I.H. Bastiaens, “Imaging biochemistry inside cells,” **Trends in Cell Biol.**, vol. 11, no. 5, pp. 203–211, May 2001.
- [2] Eric Betzid, George H. Patterson, Rachid Sougrat, O. Wolf Linwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess, “Imaging intracellular fluorescent proteins at nanometer resolution,” **Science**, vol. 313, pp. 1642–1645, September 2006.
- [3] Vasilis Ntziachristos, Jorge Ripoll, Lihong V Wang, and Ralph Weissleder, “Looking and listening to light: the evolution of whole-body photonic imaging,” **Nature Biotech.**, vol. 23, no. 3, pp. 313–320, March 2005.
- [4] Edward E. Graves, Jorge Ripoll, Ralph Weissleder, and Vasilis Ntziachristos, “A submillimeter resolution fluorescence molecular imaging system for small animal imaging,” **Med. Physics**, vol. 30, no. 5, pp. 901–911, May 2003.
- [5] Vasilis Ntziachristos, Eyk A. Schellenberger, Jorge Ripoll, Doreen Yessayan, Edward Graves, Jr. Alexi Bogdanov, Lee Josephson, and Relph Weissleder, “Visualization of antitumor treatment by means of fluorescence molecular tomography with an annexin v-cy5.5 conjugate,” **Proc Nat. Acad. Sci.**, vol. 101, no. 33, pp. 12294–12299, Aug 2004.

- [6] Anuradha Godavarty, Margaret J. Eppstein, Chaoyang Zhang, and Eva M. Sevick-Muraca, "Detection of single and multiple targets in tissue phantoms with fluorescence-enhanced optical images: Feasibility study," **Radiology**, vol. 235, no. 1, pp. 148–154, April 2005.
- [7] Jangwoen Lee and Eva M. Sevick-Muraca, "Three-dimensional fluorescence enhanced optical tomography using referenced frequency-domain photon migration measurements at emission and excitation wavelengths," **J. Opt. Soc. Am. A**, vol. 19, no. 4, pp. 759–771, April 2002.
- [8] Alexander D. Klose and Andreas H. Hielscher, "Fluorescence tomography with simulated data based on the equation of radiative transfer," **Optics Letters**, vol. 28, no. 12, pp. 1091–1021, June 2003.
- [9] Margaret J. Eppstein, Daniel J. Hawrysz, Anuradha Godavarty, and Eva M. Sevick-Muraca, "Three-dimensional, bayesian image reconstruction from sparse and noisy data sets: Near-infrared fluorescence tomography," **Proc. Nat. Acad. Sci.**, vol. 99, no. 15, pp. 9619–9624, July 2002.
- [10] Huabei Jiang, "Frequency-domain fluorescent diffusion tomograph: a finite-element-based algorithm and simulations," **Applied Optics**, vol. 37, no. 22, pp. 5337–5343, Aug 1998.
- [11] M. A. O’Leary, D. A. Boas, X. D. Li, B. Chance, and A.G. Yodh, "Fluorescence lifetime imaging in turbid media," **Optics Letters**, vol. 21, no. 2, pp. 158–160, Jan 1996.
- [12] M.J. Eppstein, F. Fedele, J. Laible, C. Zhang, A. Godavary, and E.M. Sevick-Muraca, "A comparison of exact and approximate adjoint sensitivities in fluorescence tomography," **IEEE Trans. Med. Imag.**, vol. 22, no. 10, pp. 1215–1223, Oct 2003.

- [13] Adam B. Milstein, Seungseok Oh, Kevin J. Webb, Charles A. Bouman, Quan Zhang, David A. Boas, and R.P. Millane, “Fluorescence optical diffusion tomography,” **Applied Optics**, vol. 42, no. 16, pp. 3081–3093, June 2003.
- [14] Jenghwa Chang, Harry L. Graber, and Randall L. Barbour, “Imaging of fluorescence in highly scattering media,” **IEEE Trans Biomed Eng**, vol. 44, no. 9, pp. 810–822, Sept 1997.
- [15] G. Zacharakis, H. Kambara, H. Shih, J. Ripoll, J. Grimm, Y. Saeki, R. Weissleder, and V. Ntziachristos, “Volumetric tomography of fluorescent proteins through small animals in vivo,” **Proceedings of the National Academy of Sciences of the United States of America**, vol. 102, no. 51, pp. 18252–18257, 2005.
- [16] Jan Grimm, David G. Kirsch, Stephen D. Windsor, Carla F. Bender Kim, Philip M. Santiago, Vasilis Ntziachristos, Tyler Jacks, and Ralph Weissleder, “Use of gene expression profiling to direct in vivo molecular imaging of lung cancer,” **Proc. Nat. Acad. Sci. USA**, vol. 102, no. 40, pp. 14404–14409, Oct 2005.
- [17] David E. Sosnovik, Matthias Nahrendorf, Nikolaos Deliolas, Mikhail Novikov, Elena Aikawa, Lee Josephson, Anthony Rosenzweig, Ralph Weissleder, and Vasilis Ntziachristos, “Fluorescence tomography and magnetic resonance imaging of myocardial macrophage infiltration in infarcted myocardium in vivo,” **Circulation**, March 2007.
- [18] Jan Biemond, Reginald L. Lagendik, and Russell M. Merserau, “Iterative methods for image deblurring,” **Proc. IEEE**, vol. 78, no. 5, pp. 856–883, 1990.
- [19] A. Kienle and M. S. Patterson, “Improved solutions of the steady-state and the time-resolved diffusion equations for reflectance from a semi-infinite turbid medium,” **J. Opt. Soc. Am. A**, vol. 14, no. 1, pp. 246–254, 1997.

- [20] Simon R Arridge, "Optical tomography in medical imaging," **Inverse Problems**, , no. 15, pp. R41–R93, 1999.
- [21] A. B. Milstein, J. J. Stott, S. Oh, D. A. Boas, R. P. Millane, C. A. Bouman, and K. J. Webb, "Fluorescence optical diffusion tomography using multiple-frequency data," **Journal of the Optical Society of America a-Optics Image Science and Vision**, vol. 21, no. 6, pp. 1035–1049, 2004.
- [22] Vasilis Ntziachristos and Ralph Weissleder, "Experimental three-dimensional fluorescence reconstruction of diffuse media by use of the normalized born approximation," **Optics Letters**, vol. 26, no. 12, pp. 893–895, June 2001.
- [23] Antonie Soubret, Jorge Ripoll, and Vasilis Ntziachristos, "Accuracy of fluorescent tomography in the presence of heterogeneities: Study of the normalized born ratio," **IEEE Trans. Med. Imag**, vol. 24, no. 10, pp. 1377–1386, Oct 2005.
- [24] Damon Hyde, Eric Miller, Dana H. Brooks, and Vasilis Ntziachristos, "A statistical method for inverting the born ratio," **to appear: IEEE Trans. Med Imag**, 2007.
- [25] V. Ntziachristos, A. G. Yodh, M. D. Schnall, and B. Chance, "Mri-guided diffuse optical spectroscopy of malignant and benign breast lesions," **Neoplasia**, vol. 4, pp. 347–354, 2002.
- [26] V. Ntziachristos, A. G. Yodh, M. Schnall, and B. Chance, "Concurrent mri and diffuse optical tomography of breast after indocyanine green enhancement," **Proceedings Of The National Academy Of Sciences Of The United States Of America**, vol. 97, pp. 2767–2772, 2000.
- [27] B. Brooksby, S. D. Jiang, H. Dehghani, B. W. Pogue, K. D. Paulsen, J. Weaver, C. Kogel, and S. P. 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,” **Journal of Biomedical Optics**, vol. 10, no. 5, 2005.
- [28] B. A. Brooksby, H. Dehghani, B. W. Pogue, and K. D. Paulsen, “Near-infrared (nir) tomography breast image reconstruction with a priori structural information from mri: Algorithm development for reconstructing heterogeneities,” **IEEE Journal of Selected Topics in Quantum Electronics**, vol. 9, no. 2, pp. 199–209, 2003.
- [29] G. Boverman, E. L. Miller, L. Ang, Q. Zhang, T. Chaves, D. H. Brooks, and D. A. Boas, “Quantitative spectroscopic diffuse optical tomography of the breast guided by imperfect a priori structural information,” **Physics in Medicine and Biology**, vol. 50, no. 17, pp. 3941–3956, 2005.
- [30] Scott C. Davis, Hamid Dehghani, Jia Wang, Shudong Jiang, Brian W. Pogue, and Keith D. Paulsen, “Image-guided diffuse optical fluorescence tomography implemented with laplacian-type regularization,” **Optics Express**, vol. 15, no. 7, pp. 4066–4082, April 2007.
- [31] Phaneendra K. Yalavarthy, Brian W. Pogue, Hamid Dehghani, Colin M. Carpenter, Shudong Jiang, and Keith D. Paulsen, “Structural information within regularization matrices improves near infrared diffuse optical tomography,” **Optics Express**, vol. 15, no. 13, pp. 8043–8058, June 2007.
- [32] M. Abramowitz and MW Davidson, “Molecular expressions : Introduction to microscopy,” <http://micro.magnet.fsu.edu/primer/index.html> : Retrieved Dec 1, 2008.
- [33] M. Cutler, “Transillumination as an aid in the diagnosis of breast lesions,” **Surg. Gynec. Obstet.**, vol. 48, pp. 721–728, 1929.
- [34] Kevin K. Tremper and Steven J. Barker, “Pulse oximetry,” **Anesthesiology**, vol. 70, pp. 98–108, 1989.

- [35] A. P. Gibson, J. C. Hebden, and S. R. Arridge, "Recent advances in diffuse optical imaging," **Physics in Medicine and Biology**, vol. 50, no. 4, pp. R1–R43, 2005.
- [36] Simon R Arridge and Jeremy C Hebden, "Optical imaging in medicine: Ii. modelling and reconstruction," **Phys Med Bio**, , no. 42, pp. 841–853, 1997.
- [37] David A. Boas, Dana H. Brooks, Eric L. Miller, Charles A. DiMarzio, Misha Kilmer, Richard J. Gaudette, and Quan Zhang, "Imaging the body with diffuse optical tomography," **IEEE Signal Processing Magazine**, pp. 57–75, November 2001.
- [38] Simon R. Arridge, Jeremy C. Hebden, Martin Schweiger, Florian E.W. Schmidt, Martin E Fry, Elizabeth M.C Hillman, Hamid Dehghani, and David T. Delpy, "A method for three-dimensional time-resolved optical tomography," **International Journal for Imaging Systems and Technology**, vol. 11, pp. 2–11, 2000.
- [39] Simon R Arridge, "The theoretical basis for the determination of optical path-lengths in tissue," **Phys med Bio**, , no. 37, pp. 1531–1560, 1992.
- [40] Louise C Enfield, Adam P. Gibson, Nicholas L. Everdell, David T. Delpy, Martin Schweiger, Simon R. Arridge, Caroline Richardson, Mohammad Keshtgar, Michael Douek, and Jeremy C. Hebden, "Three-dimensional time-resolved optical mammography of the uncompressed breast," **Applied Optics**, vol. 46, no. 17, pp. 3628–3638, June 2007.
- [41] J. C. Hebden, A. Gibson, T. Austin, R. M. Yusof, N. Everdell, D. T. Delpy, S. R. Arridge, J. H. Meek, and J. S. Wyatt, "Imaging changes in blood volume and oxygenation in the newborn infant brain using three-dimensional optical tomography," **Physics in Medicine and Biology**, vol. 49, no. 7, pp. 1117–1130, 2004.
- [42] A. P. Gibson, T. Austin, N. L. Everdell, M. Schweiger, S. R. Arridge, J. H. Meek, J. S. Wyatt, D. T. Delpy, and J. C. Hebden, "Three-dimensional whole-head optical

- passive motor evoked responses in the tomography of neonate,” **Neuroimage**, vol. 30, no. 2, pp. 521–528, 2006.
- [43] T. S. Leung, C. E. Elwell, and D. T. Delpy, “Estimation of cerebral oxy- and deoxy-haemoglobin concentration changes in a layered adult head model using near-infrared spectroscopy and multivariate statistical analysis,” **Physics in Medicine and Biology**, vol. 50, no. 24, pp. 5783–5798, 2005.
- [44] B. Brooksby, S. Srinivasan, S. D. Jiang, H. Dehghani, B. W. Pogue, K. D. Paulsen, J. Weaver, C. Kogel, and S. P. Poplack, “Spectral priors improve near-infrared diffuse tomography more than spatial priors,” **Optics Letters**, vol. 30, no. 15, pp. 1968–1970, 2005.
- [45] S. Srinivasan, B. W. Pogue, B. Brooksby, S. D. Jiang, H. Dehghani, C. Kogel, W. A. Wells, S. P. Poplack, and K. D. Paulsen, “Near-infrared characterization of breast tumors in vivo using spectrally-constrained reconstruction,” **Technology in Cancer Research and Treatment**, vol. 4, no. 5, pp. 513–526, 2005.
- [46] A. Corlu, R. Choe, T. Durduran, K. Lee, M. Schweiger, S. R. Arridge, E. M. C. Hillman, and A. G. Yodh, “Diffuse optical tomography with spectral constraints and wavelength optimization,” **Applied Optics**, vol. 44, no. 11, pp. 2082–2093, 2005.
- [47] C. H. Schmitz, H. L. Graber, Y. L. Pei, M. Farber, M. Stewart, R. D. Levina, M. B. Levin, Y. Xu, and R. L. Barbour, “Dynamic studies of small animals with a four-color diffuse optical tomography imager,” **Review of Scientific Instruments**, vol. 76, no. 9, 2005.
- [48] S Chandrasekhar, **Radiative Transfer**, Dover, New York, NY, 1960.

- [49] A. D. Klose and A. H. Hielscher, “Optical tomography using the time-independent equation of radiative transfer - part 2: inverse model,” **Journal of Quantitative Spectroscopy & Radiative Transfer**, vol. 72, no. 5, pp. 715–732, 2002.
- [50] A. D. Klose, U. Netz, J. Beuthan, and A. H. Hielscher, “Optical tomography using the time-independent equation of radiative transfer - part 1: forward model,” **Journal of Quantitative Spectroscopy & Radiative Transfer**, vol. 72, no. 5, pp. 691–713, 2002.
- [51] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, “The inverse source problem based on the radiative transfer equation in optical molecular imaging,” **Journal of Computational Physics**, vol. 202, no. 1, pp. 323–345, 2005.
- [52] Greg Boverman, “Modeling and nonlinear inversion for frequency domain diffuse optical tomography,” M.S. thesis, Northeastern University, Boston MA, June 2003.
- [53] John C Schotland, “Continuous-wave diffusion imaging,” **J. Opt. Soc. Am. A**, vol. 14, no. 1, pp. 275–279, 1997.
- [54] Richard C. Haskell, Lars O. Svaasand, Tsong-Tseh Tsay, Ti-Chen Feng, Matthew S McAdams, and Bruce J Tromberg, “Boundary conditions for the diffusion equation in radiative transfer,” **J. Opt. Soc. Am. A**, vol. 11, no. 10, pp. 2727–2741, October 1994.
- [55] Jorge Ripoll and Vasilis Ntziachristos, “Iterative boundary method for diffuse optical tomography,” **J. Opt. Soc. Am A**, vol. 20, no. 6, 2003.
- [56] Jorge Ripoll, Vasilis Ntziachristos, Remi Carminati, and Manuel Nieto-Vesperinas, “Kirchoff approximation for diffusive waves,” **Phys. Rev. E**, vol. 64, no. 5, 2001.
- [57] J. Ripoll and V. Ntziachristos, “From finite to infinite volumes: Removal of boundaries in diffuse wave imaging,” **Physical Review Letters**, vol. 96, no. 17, 2006.

- [58] Simon R Arridge, “Photon-measurement density function. part 1: Analytical forms,” **Applied Optics**, , no. 34, pp. 7395–7409, 1995.
- [59] Ranadhit Roy and Eva M. Sevick-Muraca, “Three-dimensional unconstrained and constrained image-reconstruction techniques applied to fluorescence, frequency-domain photon migration,” **Applied Optics**, vol. 40, no. 13, pp. 2206–2214, May 2001.
- [60] A. Joshi, W. Bangerth, K. Hwang, J. Rasmussen, and E. M. Sevick-Muraca, “Plane-wave fluorescence tomography with adaptive finite elements,” **Optics Letters**, vol. 31, no. 2, pp. 193–195, 2006.
- [61] Jae Hoon Lee, Amit Joshi, and Eva M. Sevick-Muraca, “Fully adaptive finite element based tomography using tetrahedral dual-meshing for fluorescence enhanced optical imaging in tissue,” **Optics Express**, vol. 15, no. 11, pp. 6956–6975, May 2007.
- [62] E. E. Graves, J. P. Culver, J. Ripoll, R. Weissleder, and V. Ntziachristos, “Singular-value analysis and optimization of experimental parameters in fluorescence molecular tomography,” **Journal of the Optical Society of America a-Optics Image Science and Vision**, vol. 21, no. 2, pp. 231–241, 2004.
- [63] Nikolaos Deliolanis, Tobias Lasser, Damon Hyde, Antoine Soubret, Jorge Ripoll, and Vasilis Ntziachristos, “Free-space fluorescence molecular tomography utilizing 360 geometry projections,” **Optics Letters**, vol. 32, no. 4, pp. 382, Jan. 2007.
- [64] Tobias Lasser, Nikos Deliolanis, Antoine Soubret, Jorge Ripoll, and Vasilis Ntziachristos, “360 free space Fluorescence Molecular Tomography using silhouette surface reconstruction,” in **SPIE European Conference on Biomedical Optics**, June 2007.

- [65] Tobias Lasser, Antoine Soubret, Jorge Ripoll, and Vasilis Ntziachristos, “Surface reconstruction for free-space 360 fluorescence molecular tomography and the effects of animal motion,” **IEEE Transactions on Medical Imaging**, vol. 27, no. 2, pp. 188–194, 2008.
- [66] Andrew G. Webb, **Introduction to Biomedical Imaging**, IEEE Press Series on Biomedical Engineering. Wiley - IEEE Press, 2002.
- [67] Jerrold T Bushberg, J. Anthony Seibert, Edwin M. Leidholdt Jr, and John M. Boone, **The Essential Physics of Medical Imaging**, Lippincott Williams and Wilkins, 2nd edition, 2001.
- [68] Z. M. Wang, G. Y. Panasyuk, V. A. Markel, and J. C. Schotland, “Experimental demonstration of an analytic method for image reconstruction in optical diffusion tomography with large data sets,” **Optics Letters**, vol. 30, no. 24, pp. 3338–3340, 2005.
- [69] John C. Schotland and Vadim A Markel, “Fourier-laplace structure of the inverse scattering problem for the radiative transport equation,” **Inverse Problems and Imaging**, vol. 1, no. 1, pp. 181–188, 2007.
- [70] Gene H. Golub and Charles F. Van Loan, **Matrix Computations - Third Edition**, The John Hopkins University Press, Baltimore and London, 1996.
- [71] P. C. Hansen, “Analysis of discrete ill-posed problems by means of the l-curve,” **Siam Review**, vol. 34, no. 4, pp. 561–580, 1992.
- [72] J. Kaipio and E. Somersalo, **Statistical and Computational Inverse Problems**, vol. 160 of **Applied Mathematical Sciences**, Springer Verlag, 2004.
- [73] X. Intes, V. Ntziachristos, J. P. Culver, A. Yodh, and B. Chance, “Projection access order in algebraic reconstruction technique for diffuse optical tomography,” **Physics in Medicine and Biology**, vol. 47, no. 1, pp. N1–N10, 2002.

- [74] Jonathan Richard ShewChuck, “An introduction to the conjugate gradient method without the agonizing pain,” Tech. Rep., Carnegie Mellon University, 1994.
- [75] Christopher Paige and Michael Saunders, “Lsqr: An algorithm for sparse linear equations and sparse least squares,” **ACM Transactions on Mathematical Software**, vol. 8, no. 1, pp. 43–71, 1982.
- [76] James A Nagy and Zdenek Strakos, “Enforcing nonnegativity in image reconstruction algorithms,” in **Mathematical Modeling, Estimation, and Imaging**. SPIE, 2000, vol. 4121.
- [77] A. Kak and M. Slaney, **Computerized Tomographic Imaging**, IEEE Press, 1987.
- [78] Per Christian Hansen, **Rank Deficient and Discrete Ill-Posed Problems**, SIAM Monographs on Mathematical Modeling and Computation, 1998.
- [79] Misha E. Kilmer and Dianne P. O’Leary, “Choosing regularization parameters in iterative methods for ill-posed problems,” **SIAM J. Matrix Analysis**, vol. 22, no. 4, pp. 1204–1221, 2001.
- [80] B. Saleh, **Photoelectron Statistics with Applications to Spectroscopy and Optical Communications**, Springer-Verlag, Berlin, 1978.
- [81] D. L. Snyder, **Random Point Processes**, Wiley-Interscience, New York, 1975.
- [82] Athanasios Papoulis and S. Unnikrishna Pillai, **Probability, Random Variables and Stochastic Processes**, McGraw-Hill, 4th edition, 2001.
- [83] D. V. Hinkley, “On the ratio of two correlated normal random variables,” **Biometrika**, vol. 56, no. 3, pp. 635–639, 1969.

- [84] Brian W Pogue, Troy O McBride, Judith Prewitt, Ulf L Osterberg, and Keith D Paulsen, “Spatially variant regularization improves diffuse optical tomography,” **Applied Optics**, vol. 38, no. 13, 1999.
- [85] P. K. Yalavarthy, B. W. Pogue, H. Dehghani, and K. D. 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.
- [86] M. Guven, B. Yazici, X. Intes, and B. Chance, “Diffuse optical tomography with a priori anatomical information,” **Physics in Medicine and Biology**, vol. 50, no. 12, pp. 2837–2858, 2005.
- [87] M. Schweiger and S. R. Arridge, “Optical tomographic reconstruction in a complex head model using a priori region boundary information,” **Physics in Medicine and Biology**, vol. 44, no. 11, pp. 2703–2721, 1999.
- [88] A. Li, G. Boverman, Y. H. Zhang, D. Brooks, E. L. Miller, M. E. Kilmer, Q. Zhang, E. M. C. Hillman, and D. A. Boas, “Optimal linear inverse solution with multiple priors in diffuse optical tomography,” **Applied Optics**, vol. 44, no. 10, pp. 1948–1956, 2005.
- [89] A. Li, E. L. Miller, M. E. Kilmer, T. J. Brunkilacchio, T. Chaves, J. Stott, Q. Zhang, T. Wu, M. Chorlton, R. H. Moore, D. B. Kopans, and D. A. Boas, “Tomographic optical breast imaging guided by three-dimensional mammography,” **Applied Optics**, vol. 42, no. 25, pp. 5181–5190, 2003.
- [90] D. A. Boas and A. M. Dale, “Simulation study of magnetic resonance imaging-guided cortically constrained diffuse optical tomography of human brain function,” **Applied Optics**, vol. 44, no. 10, pp. 1957–1968, 2005.
- [91] M. M. Huang, T. Q. Xie, N. G. Chen, and Q. Zhu, “Simultaneous reconstruction of absorption and scattering maps with ultrasound localization: feasibility study

- using transmission geometry,” **Applied Optics**, vol. 42, no. 19, pp. 4102–4114, 2003.
- [92] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo, “Inverse problems with structural prior information (vol 14, pg 713, 1999),” **Inverse Problems**, vol. 15, no. 4, pp. 1111–1111, 1999.
- [93] B. Kanmani and R. M. Vasu, “Diffuse optical tomography using intensity measurements and the a priori acquired regions of interest: theory and simulations,” **Physics in Medicine and Biology**, vol. 50, no. 2, pp. 247–264, 2005.
- [94] Imaging breast adipose and fibroglandular tissue molecular signatures by using hybrid MRI-guided near-infrared spectral tomography, “Ben brooksby and brian w. pogue and shudong jiang and hamid dehghani and subhadra srinivasan and christine kogel and tor d. testeson and john weaver and steven p poplack and keith d. paulsen,” **Proc. Nat. Acad. Sci. USA**, vol. 103, no. 23, pp. 8828–8833, June 2006.
- [95] X. Intes, C. Maloux, M. Guven, B. Yazici, and B. Chance, “Diffuse optical tomography with physiological and spatial a priori constraints,” **Physics in Medicine and Biology**, vol. 49, no. 12, pp. N155–N163, 2004.
- [96] James Nagy and Zdenek Strakos, “Enforcing nonnegativity in image reconstruction algorithms,” in **Mathematical Modeling, Estimation, and Imaging**, David C. Wilson, Ed., pp. 182–190. 2000.
- [97] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Q. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer, and M. Scazufca, “Global prevalence of dementia: a delphi consensus study,” **Lancet**, vol. 366, no. 9503, pp. 2112–2117, 2005.

- [98] R. L. Nussbaum and C. E. Ellis, “Genomic medicine: Alzheimer’s disease and parkinson’s disease,” **New England Journal of Medicine**, vol. 348, no. 14, pp. 1356–1364, 2003.
- [99] M. Goedert and M. G. Spillantini, “A century of alzheimer’s disease,” **Science**, vol. 314, no. 5800, pp. 777–781, 2006.
- [100] Michel Goedert, “A century of alzheimer’s disease,” **Science**, vol. 314, pp. 777–781, November 2006.
- [101] Jean Marx, “A new take on tau,” **Science**, vol. 316, pp. 1416–1417, June 2007.
- [102] K. SantaCruz, J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingelsson, A. Guimaraes, M. DeTure, M. Ramsden, E. McGowen, C. Forster, M. Yue, J. Orne, C. Janus, A. Mariash, M. Kuskowski, B. Hyman, M. Hutton, and K.H. Ashe, “Tau suppression in a neurodegenerative mouse model improves memory function,” **Science**, vol. 309, pp. 476–481, 15 July 2005.
- [103] M. A. Chishti, D. S. Yang, C. Janus, A. L. Phinney, P. Horne, J. Pearson, R. Strome, N. Zuker, J. Loukides, J. French, S. Turner, G. Lozza, M. Grilli, S. Kunkin, C. Morissette, J. Paquette, F. Gervais, C. Bergeron, P. E. Fraser, G. A. Carlson, P. St George-Hyslop, and D. Westaway, “Early-onset amyloid deposition and cognitive deficits in transgenic mice expressing a double mutant form of amyloid precursor protein 695,” **Journal of Biological Chemistry**, vol. 276, no. 24, pp. 21562–21570, 2001.
- [104] K. Hsiao, P. Chapman, S. Nilsen, C. Eckman, Y. Harigaya, S. Younkin, F. S. Yang, and G. Cole, “Correlative memory deficits, a beta elevation, and amyloid plaques in transgenic mice,” **Science**, vol. 274, no. 5284, pp. 99–102, 1996.
- [105] D. Games, D. Adams, R. Alessandrini, R. Barbour, P. Berthelette, C. Blackwell, T. Carr, J. Clemens, T. Donaldson, F. Gillespie, T. Guido, S. Hagopian, K. John-

- sonwood, K. Khan, M. Lee, P. Leibowitz, I. Lieberburg, S. Little, E. Masliah, L. McConlogue, M. Montoyazavala, L. Mucke, L. Paganini, E. Penniman, M. Power, D. Schenk, P. Seubert, B. Snyder, F. Soriano, H. Tan, J. Vitale, S. Wadsworth, B. Wolozin, and J. Zhao, "Alzheimer-type neuropathology in transgenic mice over-expressing v717f beta-amyloid precursor protein," **Nature**, vol. 373, no. 6514, pp. 523–527, 1995.
- [106] Christine Sturchler-Pierrat, Dorothee Abramowski, Mairead Duke, Karl-Heinz Wiederhold, Claudia Mistl, Sabin Rothacher, Birgit Ledermann, Kurt Burki, Peter Frey, Paolo A. Pagnetti, Caroline Waridel, Michael E. Calhoun, Mathias Jucker, Alphonse Probst, Matthias Staufenbiel, and Bernd Sommer, "Two amyloid precursor protein transgenic mouse models with alzheimer disease-like pathology," **Proc. Nat. Acad. Sci.**, vol. 94, pp. 13287–13292, 1997.
- [107] C. Sturchler-Pierrat and M. Staufenbiel, "Pathogenic mechanisms of alzheimer's disease analyzed in the app23 transgenic mouse model," **Ann N Y Acad Sci**, vol. 920, pp. 134–9, 2000.
- [108] L. Prut, D. Abramowski, T. Krucker, C. L. Levy, A. J. Roberts, M. Staufenbiel, and C. Wiessner, "Aged app23 mice show a delay in switching to the use of a strategy in the barnes maze," **Behav Brain Res**, vol. 179, no. 1, pp. 107–10, 2007.
- [109] T. Krucker, A. Schuler, E. P. Meyer, M. Staufenbiel, and N. Beckmann, "Magnetic resonance angiography and vascular corrosion casting as tools in biomedical research: application to transgenic mice modeling alzheimer's disease," **Neurol Res**, vol. 26, no. 5, pp. 507–16, 2004.
- [110] N. Beckmann, A. Schuler, T. Mueggler, E. P. Meyer, K. H. Wiederhold, M. Staufenbiel, and T. Krucker, "Age-dependent cerebrovascular abnormalities and blood flow

- disturbances in app23 mice modeling alzheimer's disease," **J Neurosci**, vol. 23, no. 24, pp. 8453–9, 2003.
- [111] Eric P. Meyer, Alexandra Ulmann-Schuler, Matthias Staufenbiel, and Thomas Krucker, "Altered morphology and 3d architecture of brain vasculature in a mouse model for alzheimer's disease," **Proc. Nat. Acad. Science USA**, vol. 105, no. 9, pp. 3587–3592, 2008.
- [112] D. A. Hammoud, J. M. Hoffman, and M. G. Pomper, "Molecular neuroimaging: From conventional to emerging technique," **Radiology**, vol. 245, no. 1, pp. 21–42, 2007.
- [113] E. D. Agdeppa, V. Kepe, K. Shoghi-Jadid, J. Liu, G. W. Small, S. C. Huang, A. Petric, N. Satyamurthy, and J. R. Barrio, "2-dialkylamino-6-acylmalononitrile substituted naphthalenes (ddnp analogs): Novel diagnostic and therapeutic tools in alzheimer's disease (vol 5, pg 404, 2003)," **Molecular Imaging and Biology**, vol. 6, no. 1, pp. 68–68, 2004.
- [114] E. D. Agdeppa, V. Kepe, J. Liu, S. Flores-Torres, N. Satyamurthy, A. Petric, G. M. Cole, G. W. Small, S. C. Huang, and J. R. Barrio, "Binding characteristics of radiofluorinated 6-dialkylamino-2-naphthylethylidene derivatives as positron emission tomography imaging probes for beta-amyloid plaques in alzheimer's disease," **Journal of Neuroscience**, vol. 21, no. 24, pp. art. no.–RC189, 2001.
- [115] G. W. Small, V. Kepe, L. M. Ercoli, P. Siddarth, S. Y. Bookheimer, K. J. Miller, H. Lavretsky, A. C. Burggren, G. M. Cole, H. V. Vinters, P. M. Thompson, S. C. Huang, N. Satyamurthy, M. E. Phelps, and J. R. Barrio, "Pet of brain amyloid and tau in mild cognitive impairment," **New England Journal of Medicine**, vol. 355, no. 25, pp. 2652–2663, 2006.

- [116] C. C. Rowe, S. Ng, U. Ackermann, S. J. Gong, K. Pike, G. Savage, T. F. Cowie, K. L. Dickinson, P. Maruff, D. Darby, C. Smith, M. Woodward, J. Merory, H. Tochon-Danguy, G. O’Keefe, W. E. Klunk, C. A. Mathis, J. C. Price, C. L. Masters, and V. L. Villemagne, “Imaging beta-amyloid burden in aging and dementia,” **Neurology**, vol. 68, no. 20, pp. 1718–1725, 2007.
- [117] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackowiak, “A voxel-based morphometric study of ageing in 465 normal adult human brains,” **Neuroimage**, vol. 14, no. 1, pp. 21–36, 2001.
- [118] Martin Hintersteiner, Albert Enz, Peter Frey, Anne-Lise Jaton, Willy Kinzy, Rainer Kneuer, Ulf Neumann, Markus Rudin, Matthias Staufenbiel, Markus Stoeckli, Karl-Heinz Wiederhold, and Hand-Ulright Gremlish, “In vivo detection of amyloid- β deposits by near-infrared imaging using an oxazine-derivative probe,” **Nature Biotechnology**, vol. 23, no. 5, pp. 577–583, May 2005.
- [119] M. Y. Han, X. H. Gao, J. Z. Su, and S. Nie, “Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules,” **Nature Biotechnology**, vol. 19, no. 7, pp. 631–635, 2001.
- [120] B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, and R. Y. Tsien, “Review - the fluorescent toolbox for assessing protein location and function,” **Science**, vol. 312, no. 5771, pp. 217–224, 2006.
- [121] E. M. Merzlyak, J. Goedhart, D. Shcherbo, M. E. Bulina, A. S. Shcheglov, A. F. Fradkov, A. Gaintzeva, K. A. Lukyanov, S. Lukyanov, T. W. J. Gadella, and D. M. Chudakov, “Bright monomeric red fluorescent protein with an extended fluorescence lifetime,” **Nature Methods**, vol. 4, no. 7, pp. 555–557, 2007.

- [122] X. K. Shu, N. C. Shaner, C. A. Yarbrough, R. Y. Tsien, and S. J. Remington, “Novel chromophores and buried charges control color in mfruits,” **Biochemistry**, vol. 45, no. 32, pp. 9639–9647, 2006.
- [123] D. Shcherbo, E. M. Merzlyak, T. V. Chepurnykh, A. F. Fradkov, G. V. Ermakova, E. A. Solovieva, K. A. Lukyanov, E. A. Bogdanova, A. G. Zarausky, S. Lukyanov, and D. M. Chudakov, “Bright far-red fluorescent protein for whole-body imaging,” **Nature Methods**, vol. 4, no. 9, pp. 741–746, 2007.
- [124] Vasilis Ntziachristos, Ching-Hsuan Tung, Christoph Bremer, and Ralph Weissleder, “Fluorescence molecular tomography resolves protease activity in vivo,” **Nature Medicine**, vol. 8, no. 7, pp. 757–761, 2002.
- [125] A. Villringer and B. Chance, “Non-invasive optical spectroscopy and imaging of human brain function,” **Trends in Neurosciences**, vol. 20, no. 10, pp. 435–442, 1997.
- [126] K. Tanner, E. D’Amico, A. Kaczmarowski, S. Kukreti, J. Malpeli, W. W. Mantulin, and E. Gratton, “Spectrally resolved neurophotonic: a case report of hemodynamics and vascular components in the mammalian brain,” **Journal of Biomedical Optics**, vol. 10, no. 6, 2005.
- [127] M. Vauhkonen, D. Vadasz, P. A. Karjalainen, E. Somersalo, and J. P. Kaipio, “Tikhonov regularization and prior information in electrical impedance tomography,” **IEEE Trans. Med. Imag.**, vol. 17, no. 2, pp. 285–293, April 1998.
- [128] Daniel R. Lynch and Dennis J. McGillicuddy Jr., “Objective analysis for coastal regimes,” **Continental Shelf Research**, vol. 21, pp. 1299–1315, 2001.
- [129] Emmanuel Candes, Justin Romberg, and Terence Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” **IEEE Trans. Inf. Theory**, vol. 52, no. 2, pp. 489–509, 2006.

- [130] David Donoho, “Compressed sensing,” **IEEE Trans Inf. Theory**, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [131] Michael Grant and Stephen Boyd, “Cvx: Matlab software for disciplined convex programming (web page and software),” <http://stanford.edu/~boyd/cvx>, October 2008.
- [132] Michael Grant and Stephen Boyd, **Recent Advances in Learning and Control (a tribute to M. Vidyasagar)**, chapter Graph implementations for nonsmooth convex programs, Lecture Notes in Control and Information Sciences. Springer, http://stanford.edu/~boyd/graph_dcp.html, 2008.
- [133] Loren T Niklason, Bradley T Christian, Laura E Niklason, Daniel B Kopans, Donald E Castleberry, Beale H Opsahl-Ong, Cynthia E Landberg, Priscilla J Slanetz, Angela A Giardino, Richard Moore, Douglas Albagli, Michael C DeJule, Paul F Fitzgerald, David F Fobare, Brian W. Giambattista, Robert F Kwasnick, Jianqiang Liu, Stanley J. Lubowski, George E. Possin, James F. Richotte, Ching-Yen Wei, and Reinhold F. Wirth, “Digital tomosynthesis in breast imaging,” **Radiology**, vol. 205, pp. 399–406, 1997.
- [134] J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane, “Nonlinear multigrid algorithms for bayesian optical diffusion tomography,” **IEEE Transactions on Image Processing**, vol. 10, no. 6, pp. 909–922, 2001.

Appendix A

Detailed Derivations of Density Approximation

A.1 Approximation of terms involving Φ

In section ??, our first step in simplifying the statistical density for the data was to make the approximation:

$$\left[\Phi \left\{ \frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \right\} - \Phi \left\{ -\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \right\} \right] \simeq 1 \quad (\text{A.1})$$

in order to eliminate the integrals from the density. We show here the conditions under which this approximation is valid. Clearly, as $\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} \rightarrow \infty$, (A.1) is valid, and equivalently the integral:

$$\int_0^{b(\mathbf{u}_i)/a(\mathbf{u}_i)} \phi(u) du \quad (\text{A.2})$$

is equal to 0.5. Using the definition of $\text{erf}()$:

$$\int_0^q \phi(t) dt = \frac{1}{2} \text{erf} \left(\frac{q}{\sqrt{2}} \right) \quad (\text{A.3})$$

we can write the inequality:

$$0.5 - \frac{1}{2} \text{erf} \left(\frac{b(\mathbf{u}_i)}{\sqrt{2}a(\mathbf{u}_i)} \right) < \frac{\epsilon}{2} \quad (\text{A.4})$$

where we define ϵ as the tolerance allowed in the final approximation. From this we can obtain the values for the ratio $\frac{a(\mathbf{u}_i)}{b(\mathbf{u}_i)}$ which satisfy the inequality. We rearrange the inequality to obtain:

$$\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} > \sqrt{2} \text{erf}^{-1}(1 - \epsilon) = \delta \quad (\text{A.5})$$

where δ is set equal to the bound. Setting $\epsilon = 2 * 10^{-5}$ in (A.4) yields $\delta = 4.42$ as the required condition for (A.1) to be an appropriate approximation. Even increasing the precision to $\epsilon = 2 * 10^{-16}$ only requires $\delta = 8.21$.

However, $a(\mathbf{u}_i)$ and $b(\mathbf{u}_i)$ are compound quantities which depend on the physical quantities $m_{fl,i}$ and $m_{exc,i}$ in a complicated fashion (See (3.9)). Thus we want to know equivalent bounds on the values of $m_{exc,i}$ and $m_{fl,i}$, given a specified value for δ . We start with:

$$\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} = \frac{\frac{m_{fl,i}\mathbf{u}_i}{\sigma^2(m_{fl,i})} + \frac{m_{exc,i}}{\sigma^2(m_{exc,i})}}{\left(\frac{\mathbf{u}_i^2}{\sigma^2(m_{fl,i})} + \frac{1}{\sigma^2(m_{exc,i})}\right)^{1/2}}. \quad (\text{A.6})$$

By placing this into the equation $\frac{b(\mathbf{u}_i)}{a(\mathbf{u}_i)} > \delta$, squaring both sides, and rearranging, we can obtain:

$$\begin{aligned} \sigma^2(m_{exc,i})\mathbf{u}_i^2 \left(\frac{m_{fl,i}^2}{\sigma^2(m_{fl,i})} - \delta^2 \right) + \\ \sigma^2(m_{fl,i}) \left(\frac{m_{exc,i}^2}{\sigma^2(m_{exc,i})} - \delta^2 \right) > -2m_{fl,i}\mathbf{u}_im_{exc,i} \end{aligned} \quad (\text{A.7})$$

When both $\frac{m_{fl,i}}{\sigma(m_{fl,i})}$ and $\frac{m_{exc,i}}{\sigma(m_{exc,i})}$ are greater than δ , the inequality is satisfied because the left hand side will be positive while the right hand side will be negative (given that \mathbf{u}_i , $m_{exc,i}$ and $m_{fl,i}$ are all positive quantities). For $\delta = 4.42$, this occurs when both $m_{exc,i}$ and $m_{fl,i}$ are greater than 18 counts. In general, however, we want to be able to use as much of our data as possible, so we consider the case where one of the two values is less than δ . This will result in one of the two left hand terms being negative rather than positive. Using the approximation $\mathbf{u}_i \simeq \frac{m_{fl,i}}{m_{exc,i}}$, we can rewrite the inequality as:

$$\begin{aligned} \frac{\sigma^2(m_{exc,i})}{m_{exc,i}^2} \left(\frac{m_{fl,i}^2}{\sigma^2(m_{fl,i})} - \delta^2 \right) + \\ \frac{\sigma^2(m_{fl,i})}{m_{fl,i}^2} \left(\frac{m_{exc,i}^2}{\sigma^2(m_{exc,i})} - \delta^2 \right) > -2. \end{aligned} \quad (\text{A.8})$$

Given that the means and variances are all positive quantities, it is easily seen that the values of the quantities inside the parentheses must be greater than $-\delta^2$. If, as assumed,

only one of the two quantities is greater than zero (ie - the associated signal is higher than 18), then it can quickly be proven that the inequality must be satisfied. Because of this, we now choose to utilize all data points where at least one of the two signals is greater than 18.

A.2 Elimination of Low Magnitude Term

Equation (3.6) in Section ?? gives the density of the data as the sum of two terms. However, it turns out that the second term is significantly smaller than the first, given reasonable conditions, thereby greatly simplifying the log-likelihood function. These conditions turn out to be achievable by minor additional requirements on the values of $m_{exc,i}$ and $m_{fl,i}$ compared to those obtained in Appendix A.1.

In the course of these derivations, we will make two primary assumptions:

$$(1) \ 0 \leq \mathbf{u}_i \leq 1.5$$

The lower end of this bound is a result of both received signals inherently being positive quantities, while the upper bound arises from analysis of many data sets. In all data sets examined, the inequality $m_{fl,i} < 1.5m_{exc,i}$ is satisfied by all data points. Thus, when taking the Born ratio, $\mathbf{u}_i = \frac{m_{fl,i}}{m_{exc,i}}$, the inequality $\mathbf{u}_i < 1.5$ will be satisfied.

$$(2) \ \mathbf{W}_i \mathbf{x} = \frac{m_{exc,i}}{m_{fl,i}} \simeq \mathbf{u}_i$$

The equality portion of this statement is the same assumption that was made in relating our statistical model to our forward model. We simply restate it here, and further assume that the value of \mathbf{x} which corresponds to the actual distribution of fluorochrome, will result in a value of $\mathbf{W}\mathbf{x}$ which closely matches the received data \mathbf{u} .

What we explicitly need to show is that the term:

$$\frac{1}{\pi \sigma(m_{fl,i}) \sigma(m_{exc,i})} e^{-c/2} \tag{A.9}$$

is small enough to be dropped from the density. We assume that it can be removed when it is several orders of magnitude smaller than the other term in the density. We take the ratio of the two terms, and study the conditions under which it will be much larger than one:

$$\frac{\pi b(\mathbf{u}_i) d(\mathbf{u}_i)}{\sqrt{2\pi} a(\mathbf{u}_i)} e^{c/2} \gg 1. \quad (\text{A.10})$$

We specify the explicit bound at which (A.10) is satisfied to be α , take the log of both sides and rearrange to get:

$$\log(d(\mathbf{u}_i)) + c/2 \geq \beta + \log(a(\mathbf{u}_i)) - \log(b(\mathbf{u}_i)) \quad (\text{A.11})$$

with:

$$\beta = \log(\alpha) + 0.5 \log(\pi/2). \quad (\text{A.12})$$

Now, given that:

$$c/2 = 0.5 \left(\frac{m_{fl,i}^2}{\sigma^2(m_{fl,i})} + \frac{m_{exc,i}^2}{\sigma^2(m_{exc,i})} \right) \quad (\text{A.13})$$

and:

$$\log(d(\mathbf{u}_i)) = \left\{ -0.5 \frac{m_{exc,i}^2 \left(\frac{m_{fl,i}}{m_{exc,i}} - \mathbf{u}_i \right)^2}{(\sigma^2(m_{exc,i}) \mathbf{u}_i + \sigma^2(m_{fl,i}))} \right\} \quad (\text{A.14})$$

we can now move the first term of $c/2$ to the right hand side and combine the second term with $\log(d(\mathbf{u}_i))$ in order to obtain:

$$0.5 \left[\frac{m_{exc,i}^2}{\sigma(m_{exc,i})^2} \left(1 - \frac{(\frac{m_{fl,i}}{m_{exc,i}} - \mathbf{u}_i)^2}{\mathbf{u}_i + \frac{\sigma^2(m_{fl,i})}{\sigma^2(m_{exc,i})}} \right) \right] \geq \quad (\text{A.15})$$

$$\beta + \log(a(\mathbf{u}_i)) - \log(b(\mathbf{u}_i)) - 0.5 \frac{m_{fl,i}^2}{\sigma(m_{fl,i})^2} \quad (\text{A.16})$$

Rearranging things a bit yields:

$$\left(1 - \frac{(\frac{m_{fl,i}}{m_{exc,i}} - \mathbf{u}_i)^2}{\mathbf{u}_i + \frac{\sigma^2(m_{fl,i})}{\sigma^2(m_{exc,i})}} \right) \geq \frac{\sigma(m_{exc,i})^2}{m_{exc,i}^2} \times \left[2\beta + 2[\log(a(\mathbf{u}_i)) - \log(b(\mathbf{u}_i))] - \frac{m_{fl,i}^2}{\sigma(m_{fl,i})^2} \right] \quad (\text{A.17})$$

Looking at the terms on the right hand side, we want to find an upper bound for $a(\mathbf{u}_i)$ and a lower bound for $b(\mathbf{u}_i)$, in order to maximize the overall value of the right hand side. Looking at $\log(b(\mathbf{u}_i))$, we have:

$$\log(b(\mathbf{u}_i)) = \log\left(\frac{m_{fl,i}\mathbf{u}_i}{\sigma^2(m_{fl,i})} + \frac{m_{exc,i}}{\sigma^2(m_{exc,i})}\right) \quad (\text{A.18})$$

$$\geq \log(m_{exc,i}) - \log(\sigma^2(m_{exc,i})) \quad (\text{A.19})$$

where we have used the lower bound on \mathbf{u}_i to establish the inequality.

For the term $a(\mathbf{u}_i)$, we want to determine an upper bound. We use our assumed upper bound on \mathbf{u}_i to obtain the inequality:

$$\log(a(\mathbf{u}_i)) \leq 0.5 \log\left(\frac{1.5^2}{\sigma^2(m_{fl,i})} + \frac{1}{\sigma^2(m_{exc,i})}\right) \quad (\text{A.20})$$

By substituting these bounds in place of $\log(b(\mathbf{u}_i))$ and $\log(a(\mathbf{u}_i))$ in (A.17) and rearranging, we get:

$$\begin{aligned} \left(\frac{m_{fl,i}}{m_{exc,i}} - \mathbf{u}_i\right)^2 &\leq \frac{\sigma^2(m_{fl,i})}{\sigma^2(m_{exc,i})} \left[1 - \frac{\sigma^2(m_{exc,i})}{m_{exc,i}^2} \times \right. \\ &\quad \left[2\beta + \log\left(\frac{1.5^2}{\sigma^2(m_{fl,i})} + \frac{1}{\sigma^2(m_{exc,i})}\right) - \right. \\ &\quad \left. \left. 2\left(\log(m_{exc,i}) - \log(\sigma^2(m_{exc,i}))\right) - \frac{m_{fl,i}^2}{\sigma^2(m_{fl,i})}\right]\right] \end{aligned} \quad (\text{A.21})$$

This puts the left hand side in terms of the mismatch between the estimate and the collected data, and makes the right hand side independent of \mathbf{u}_i . Now, we use our assumption that $\frac{m_{exc,i}}{m_{fl,i}} \simeq \mathbf{u}_i$ in order to assume that the left hand side of the inequality is less than or equal 1. This allows for a significant mismatch between the received data and our model, and provides a convenient threshold for determining a bound on which data pairs are usable. We establish this bound in a graphical fashion by simply computing the right hand side for a range of $m_{fl,i}$ and $m_{exc,i}$ values and plotting to determine where the value is greater than one. The resulting surface plot is shown in Figure A.1. From this plot, we established that all points with $m_{fl,i} > 15$ or $m_{exc,i} > 20$ will satisfy the

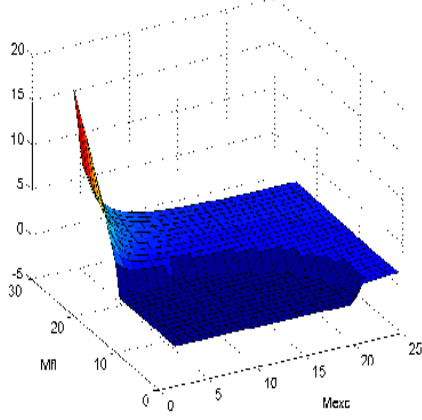


Figure A.1: Plot of the right hand side of (A.21). Those points with magnitude less than one have been set equal to -1 to give a better idea of the shape of the region in question.

criterion. Combining this requirement with the previously established requirement that one of the two signals be greater than 18 counts, we determine that all data point pairs with either $m_{fl,i} > 18$ or $m_{exc,i} > 20$ will be retained for use in the reconstruction.

A.3 Deriving the Gradient of the Cost Function

A.3.1 Primary Derivations

In the course of obtaining a maximum likelihood solution in Section ??, we derived the cost function:

$$f(\mathbf{u}, \mathbf{x}) = \sum_i g_i(x) - 0.5 \|\mathbf{Q}(\mathbf{u}, \mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2 \quad (\text{A.22})$$

which we want to maximize in order to obtain our solution. In order to simplify the notation involved in finding the gradient of (A.22), we write:

$$s(\mathbf{u}, \mathbf{x}) = 0.5 \|\mathbf{Q}(\mathbf{u}, \mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u})\|_2^2. \quad (\text{A.23})$$

We then write an equation for the gradient as:

$$\frac{df(\mathbf{u}, \mathbf{x})}{d\mathbf{x}} = \left[\sum_i \frac{dg_i(\mathbf{u}, \mathbf{x})}{d\mathbf{x}} \right] - \frac{ds(\mathbf{u}, \mathbf{x})}{d\mathbf{x}}. \quad (\text{A.24})$$

Given the complexity of the $g_i(\mathbf{u}, \mathbf{x})$ terms, we derive the gradient of those components separately in Appendix A.3.2, and make use of those results at the end of this section. Turning our attention to the second term of the gradient, we can write that gradient as:

$$\begin{aligned} \frac{ds(\mathbf{u}, \mathbf{x})}{d\mathbf{x}} = & \left[\frac{d(\mathbf{W}\mathbf{x} - \mathbf{u})}{d\mathbf{x}} \right]^T \frac{ds(\mathbf{u}, \mathbf{x})}{d(\mathbf{W}\mathbf{x} - \mathbf{u})} + \left[\frac{d\mathbf{Q}(\mathbf{u}, \mathbf{x})}{d\mathbf{x}} \right]^T \frac{ds(\mathbf{x})}{d\mathbf{Q}(\mathbf{u}, \mathbf{x})} \end{aligned} \quad (\text{A.25})$$

The first term of the sum can then be computed as:

$$\left[\frac{d(\mathbf{W}\mathbf{x} - \mathbf{u})}{d\mathbf{x}} \right]^T \frac{ds(\mathbf{x})}{d(\mathbf{W}\mathbf{x} - \mathbf{u})} = 2\mathbf{W}^T C(\mathbf{x})^T C(\mathbf{x})(\mathbf{W}\mathbf{x} - \mathbf{u}). \quad (\text{A.26})$$

To compute the second term, we use the fact that $\mathbf{q}(\mathbf{u}, \mathbf{x}) = \text{diag}(\mathbf{q}(\mathbf{u}, \mathbf{x}))$ in order to rewrite $s(\mathbf{x})$ as:

$$s(\mathbf{u}, \mathbf{x}) = \mathbf{q}(\mathbf{u}, \mathbf{x})^T \text{diag}(\mathbf{W}\mathbf{x} - \mathbf{u})^T \text{diag}(\mathbf{W}\mathbf{x} - \mathbf{u}) \mathbf{q}(\mathbf{x}). \quad (\text{A.27})$$

This rearrangement allows us to write the gradient as:

$$\frac{ds(\mathbf{u}, \mathbf{x})}{d\mathbf{q}(\mathbf{x})} = 2\text{diag}(\mathbf{W}\mathbf{x} - \mathbf{u})^T \text{diag}(\mathbf{W}\mathbf{x} - \mathbf{u}) \mathbf{q}(\mathbf{x}). \quad (\text{A.28})$$

For notational simplicity, we now define:

$$A(\mathbf{x}) = \text{diag}(\mathbf{W}\mathbf{x} - \mathbf{u}) \quad (\text{A.29})$$

and rewrite the gradient as:

$$\frac{ds(\mathbf{u}, \mathbf{x})}{d\mathbf{Q}(\mathbf{x})} = 2A(\mathbf{x})^T A(\mathbf{x}) \mathbf{q}(\mathbf{x}). \quad (\text{A.30})$$

Now we derive the value of $\frac{d\mathbf{q}(\mathbf{x})}{d\mathbf{x}}$. Given that:

$$\mathbf{q}_i(\mathbf{x}) = m_{exc,i} \left(\beta m_{exc,i} (\mathbf{u}_i + W_i \mathbf{x}) + (1 + \mathbf{u}_i) \sigma_r^2 \right)^{-1/2} \quad (\text{A.31})$$

we can write the derivative as:

$$\begin{aligned} \frac{d\mathbf{q}_i(\mathbf{x})}{d\mathbf{x}} = & \frac{-m_{exc,i}}{2} \left(\beta m_{exc,i} (\mathbf{u}_i + W_i \mathbf{x}) + (1 + \mathbf{u}_i) \sigma_r^2 \right)^{-3/2} \beta m_{exc,i} W_i \end{aligned} \quad (\text{A.32})$$

This gives us a vector, representing the gradient of a particular $\mathbf{q}_i(\mathbf{x})$ with respect to \mathbf{x} . Obtaining one for each $\mathbf{q}_i(\mathbf{x})$, we can then arrange all of these vectors into a matrix as:

$$\frac{d\mathbf{q}(\mathbf{x})}{d\mathbf{x}} = \left[\frac{d\mathbf{q}_1(\mathbf{x})^T}{d\mathbf{x}} \frac{d\mathbf{q}_2(\mathbf{x})^T}{d\mathbf{x}} \cdots \frac{d\mathbf{q}_N(\mathbf{x})^T}{d\mathbf{x}} \right] \quad (\text{A.33})$$

This, combined with the gradient of $g_i(\mathbf{x})$, derived in Appendix A.3.2, allows us to write the full gradient of the cost function as:

$$\begin{aligned} \frac{df(\mathbf{u}, \mathbf{x})}{d\mathbf{x}} = & \left[\sum_i \frac{dg_i(\mathbf{u}, \mathbf{x})}{d\mathbf{x}} \right] - \mathbf{W}^T \mathbf{Q}(\mathbf{x})^T \mathbf{Q}(\mathbf{x}) (\mathbf{W}\mathbf{x} - \mathbf{u}) - \\ & \left(\frac{d\mathbf{q}(\mathbf{x})}{d\mathbf{x}} \right)^T A(\mathbf{x})^T A(\mathbf{x}) \mathbf{q}(\mathbf{x}) \end{aligned} \quad (\text{A.34})$$

which we use to derive our fixed point iteration in the text.

A.3.2 Derivation of one of the gradient components

In this section, we explicitly determine the gradient of the $g_i(\mathbf{u}_i, \mathbf{x})$ terms. We repeat (3.19) to begin:

$$g_i(\mathbf{u}_i, \mathbf{x}) = \log(b(\mathbf{u}_i, \mathbf{x})) - \log(\sigma(m_{fl,i}(\mathbf{x}))) - 3 \log(a(\mathbf{u}_i, \mathbf{x})). \quad (\text{A.35})$$

We explicitly write the gradient of this equation:

$$\begin{aligned} \frac{dg_i(\mathbf{x})}{d\mathbf{x}} = & \frac{1}{b(\mathbf{u}_i, \mathbf{x})} \frac{db(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} \\ & - \frac{1}{\sigma(m_{fl,i}(\mathbf{x}))} \frac{d\sigma(m_{fl,i}(\mathbf{x}))}{d\mathbf{x}} - 3 \frac{1}{a(\mathbf{u}_i, \mathbf{x})} \frac{da(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}}. \end{aligned} \quad (\text{A.36})$$

This leaves us with a gradient which is the sum of three terms, each of which consists of the gradient and the reciprocals of known functions. We treat each term separately.

First we write the gradient of $b(\mathbf{u}_i, \mathbf{x})$ with respect to \mathbf{x} as:

$$\begin{aligned} \frac{db(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} = & \frac{db(\mathbf{u}_i, \mathbf{x})}{dm_{fl,i}(\mathbf{x})} \frac{dm_{fl,i}(\mathbf{x})}{d\mathbf{x}} = \\ & \left[\frac{\mathbf{u}_i \sigma(m_{fl,i}(\mathbf{x}))^2 - \beta m_{fl,i}(\mathbf{x}) \mathbf{u}_i}{\sigma(m_{fl,i}(\mathbf{x}))^4} \right] [m_{exc,i} W_{ij}]. \end{aligned} \quad (\text{A.37})$$

Combining this with (3.9) yields:

$$\begin{aligned} \frac{1}{b(\mathbf{u}_i, \mathbf{x})} \frac{db(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} = \\ \frac{(m_{exc,i} W_{ij}) \sigma(m_{exc,i})^2 [\mathbf{u}_i \sigma(m_{fl,i}(\mathbf{x})) - \beta m_{fl,i}(\mathbf{x}) \mathbf{u}_i]}{\sigma(m_{fl,i}(\mathbf{x}))^2 [\sigma(m_{exc,i})^2 m_{fl,i}(\mathbf{x}) \mathbf{u}_i + \sigma(m_{fl,i}(\mathbf{x}))^2 m_{exc,i}]} \end{aligned} \quad (\text{A.38})$$

Next, given the definition for $\sigma(m)$ in (3.3), we can write the second term in the gradient of $g_i(\mathbf{u}_i, \mathbf{x})$ as:

$$\frac{1}{\sigma(m_{fl,i}(\mathbf{x}))} \frac{d\sigma(m_{fl,i})}{d\mathbf{x}} = \frac{1}{2} \frac{\beta m_{exc,i} W_i}{\beta m_{exc,i} W_i \mathbf{x} + \sigma_r^2} \quad (\text{A.39})$$

Finally, we look at the third term, which depends on $a(\mathbf{u}_i, \mathbf{x})$, and contains the gradient:

$$\begin{aligned} \frac{da(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} = \frac{da(\mathbf{u}_i, \mathbf{x})}{dm_{fl,i}(\mathbf{x})} \frac{dm_{fl,i}(\mathbf{x})}{d\mathbf{x}} = \\ \left[\frac{-1}{2} \left(\frac{\mathbf{u}_i^2}{\sigma(m_{fl,i}(\mathbf{x}))^2} + \frac{1}{\sigma(m_{exc,i})^2} \right)^{-1/2} \times \right. \\ \left. \frac{\beta \mathbf{u}_i^2}{\sigma(m_{fl,i}(\mathbf{x}))^4} \right] [m_{exc,i} W_{ij}]. \end{aligned} \quad (\text{A.40})$$

Thus, again using (3.9), we get:

$$\frac{1}{a(\mathbf{u}_i, \mathbf{x})} \frac{da(\mathbf{u}_i, \mathbf{x})}{d\mathbf{x}} = \left[\frac{1}{2} \frac{1}{a(\mathbf{u}_i, \mathbf{x})^2} \frac{\beta \mathbf{u}_i^2}{\sigma(m_{fl,i}(\mathbf{x}))^4} \right] [m_{exc,i} W_{ij}] \quad (\text{A.41})$$

The values of (A.38), (A.41), and (A.39) allow us to write the full gradient of $g_i(\mathbf{u}_i, \mathbf{x})$.

With that in hand, we are able to explicitly evaluate the full gradient of our cost function.