

Computational Methods For The Diagnosis of Rheumatoid Arthritis With Diffuse
Optical Tomography

Ludguier D. Montejo

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

© 2014
Ludguier D. Montejo
All rights reserved

ABSTRACT

Computational Methods For The Diagnosis of Rheumatoid Arthritis With Diffuse Optical Tomography

Ludguier D. Montejo

Diffuse optical tomography (DOT) is an imaging technique where near infrared (NIR) photons are used to probe biological tissue. DOT allows for the recovery of three-dimensional maps of tissue optical properties, such as tissue absorption and scattering coefficients. The application of DOT as a tool to aid in the diagnosis of rheumatoid arthritis (RA) is explored in this work. Algorithms for improving the image reconstruction process and for enhancing the clinical value of DOT images are presented in detail. The clinical data considered in this work consists of 99 fingers from subjects with RA and 120 fingers from healthy subjects. DOT scans of the proximal interphalangeal (PIP) joint of each finger is performed with modulation frequencies of 0, 300, and 600 MHz.

A computer-aided diagnosis (CAD) framework for extracting heuristic features from DOT images and a method for using these same features to classify each joint as affected or not affected by RA is presented. The framework is applied to the clinical data and results are discussed in detail. Then, an algorithm for recovering the optical properties of biological media using the simplified spherical harmonics (SP_N) light propagation model is presented. The computational performance of the algorithm is analyzed and reported. Finally, the SP_N reconstruction algorithm is applied to clinical data of subjects with RA and the resulting images are analyzed with the CAD framework.

As the first part of the CAD framework, heuristic image features are extracted from the absorption and the scattering coefficient images using multiple compression and dimensionality reduction techniques. Overall, 594 features are extracted from the images of each joint. Then, machine learning techniques are used to evaluate the ability to discriminate between images of joints with RA

and images of healthy joints. An evolution-strategy optimization algorithm is developed to evaluate the classification strength of each feature and to find the multidimensional feature combination that results in optimal classification accuracy. Classification is performed with k -nearest neighbors (KNN), linear (LDA) and quadratic discriminate analysis (QDA), self-organizing maps (SOM), or support vector machines (SVM). Classification accuracy is evaluated based on diagnostic sensitivity and specificity values.

Strong evidence is presented that suggest there are clear difference between the tissue optical parameters of joints with RA and joints without RA. It is first shown that data obtained at 600 MHz leads to better classification results than data obtained at 300 and 0 MHz. Analysis of each extracted feature shows that DOT images of subjects with RA are statistically different ($p < 0.05$) from images of subjects without RA for over 90% of the features. Evidence shows that subjects with RA that do not have detectable signs of erosion, effusion, or synovitis (i.e. asymptomatic subjects) in MRI and US images have optical profiles similar to subjects that do have signs of erosion, effusion, or synovitis; furthermore, both of these cohorts differ from healthy controls subjects. This shows that it may be possible to accurately identify asymptomatic subjects with DOT scans. In contrast, these subjects remain difficult to identify from MRI and US images. The implications of these results are profound, as they suggest it may be possible to identify RA with DOT at an earlier stage compared to standard imaging techniques.

Results from the feature-selection algorithm show that the SVM algorithm (with a third order polynomial kernel) achieves 100.0% sensitivity and 97.8% specificity. Lower bounds for these results (at 95.0% confidence level) are 96.4% and 93.8%, respectively. Image features most predictive of RA are from the spatial variation of optical properties and the absolute range in feature values. The optimal classifiers are low dimensional combinations (< 7 features). Robust cross-validation is performed to ensure the generalization of these classification results.

The SP_N -based reconstruction algorithm uses a reduced-Hessian sequential quadratic programming (rSQP) PDE-constrained optimization approach to maximize computation efficiency. The complex-valued forward model, or frequency domain SP_N equations ($N = 1, 3$), is discretized

using the finite-volume method and solved on unstructured computational grids using the restarted GMRES algorithm. The image reconstruction algorithm is presented in detail and its performance benchmarked against the ERT algorithm. The algorithm is subsequently used to recover the absorption and scattering coefficient images of joints scanned in the RA clinical study.

While the SP_N model is inherently less accurate than the ERT model, it is nevertheless shown that the images obtained with the SP_3 -based reconstruction algorithm are sufficiently accurate and allow for the diagnosis of RA at clinically relevant sensitivity [87.9% (78.1%, 100.0%)] and specificity [92.9% (84.6%, 100.0%)] values (the 95.0% confidence interval is specified in brackets). In contrast to results obtained with the SP_3 model, the images generated with the SP_1 algorithm yield significantly lower sensitivity [66.7% (46.6%, 100.0%)] and specificity [81.0% (64.8%, 100.0%)] values. While some numerical accuracy is sacrificed by selecting the SP_3 model over the ERT model, the superior computational performance of the SP_3 algorithm allows for computation of the absorption and the scattering coefficient images in under 15 minutes and requires less than 200 MB of RAM per finger (compared to the over 180 minutes and over 6 GB of RAM needed by the ERT-based algorithm).

Overall, results indicate that the SP_3 -based reconstruction algorithm provides computational advantages over the ERT-based algorithm without sacrificing significant classification accuracy. In contrast, the SP_1 model provides computational advantages compared to the ERT at the expense of classification accuracy. This indicates that the frequency-domain SP_3 model is an ideal light propagation model for use in DOT scanning of finger joints with RA.

Altogether, the results presented in this dissertation underscore the high potential for DOT to become a clinically useful diagnostic tool. The algorithms and framework developed as part of this dissertation can be directly used on future data to help further validate the hypotheses presented in this work and to further establish DOT imaging as a valuable diagnostic tool.

Contents

List of Tables	vi
List of Figures	ix
1 Introduction	1
2 Diffuse Optical Tomography	8
2.1 Absorption and scattering of light in tissue	8
2.2 Models of light propagation in tissue	11
2.3 Equation of radiative transfer	16
2.3.1 Time-dependent ERT equation	16
2.3.2 Frequency-dependent ERT equation	18
2.3.3 Time-independent ERT equation	19
2.3.4 Numerical challenges associated with the ERT	19
2.4 Diffusion approximation to the ERT	21
2.5 SP_N approximation to the ERT	23
2.5.1 Time-independent SP_N equations	24
2.5.2 Time-dependent SP_N equations	26
2.5.3 Frequency-dependent SP_N equations	27
2.6 Structured and unstructured computational grids	30
2.7 Discussion	31
3 Rheumatoid Arthritis	33
3.1 Epidemiology	33
3.2 Diagnosis	36
3.3 Treatment	38
3.4 The role of laboratory studies	40
3.5 Optical imaging for diagnosis and monitoring	42
3.6 Discussion	44
4 Computer-Aided Diagnosis	46
4.1 Traditional Medical Imaging Modalities	47
4.2 Optical Imaging	48

4.3	Discussion	51
5	DOT Imaging Of Joints With RA	52
5.1	Background	53
5.2	Clinical study design	54
5.2.1	Composition of clinical study	54
5.2.2	Gold standard	55
5.3	Instrumentation and imaging protocol	59
5.3.1	DOT imaging instrument	59
5.3.2	DOT imaging protocol	61
5.3.3	DOT image reconstruction	63
5.3.4	Diagnostic terminology	67
5.3.5	ROC curve analysis	69
5.4	Data analysis	70
5.4.1	Region of interest and feature extraction	70
5.4.2	ANOVA and student-t test	72
5.4.3	Image classification	73
5.4.4	Modulation frequency analysis	74
5.5	Results	74
5.5.1	ANOVA analysis	76
5.5.2	Student-t test	80
5.5.3	ROC curve	80
5.5.4	LDA classification	82
5.5.5	Influence of source-modulation frequency	83
5.6	Discussion	84
6	Feature Extraction	87
6.1	Clinical data	88
6.2	Data pre-processing	89
6.3	Feature extraction	93
6.3.1	Basic features	93
6.3.2	GMM coefficients	94
6.3.3	FFT coefficients	96
6.3.4	Short-hand notation	98
6.4	Statistical analysis	99
6.4.1	Kruskal-Wallis ANOVA test and Dunn's test	99
6.4.2	Effective sample size	100
6.4.3	ROC analysis	102
6.5	Results	102
6.5.1	GMM and FFT parameterization	102
6.5.2	Kruskal-Wallis ANOVA test	103
6.5.3	Dunn's test	106
6.5.4	ROC analysis	109
6.6	Discussion	112

7	Image Classification	114
7.1	Data	116
7.2	Methods	118
7.2.1	The KNN method	118
7.2.2	The LDA and QDA methods	119
7.2.3	The SOM method	121
7.2.4	The SVM method	121
7.2.5	Cross validation and quantification of classification accuracy	129
7.2.6	Selection of best features	129
7.3	Results	132
7.3.1	KNN	136
7.3.2	LDA and QDA	136
7.3.3	SOM	137
7.3.4	SVM	138
7.3.5	Best features	139
7.4	Discussion	140
8	Finite volume SP_N Model: Forward Problem	144
8.1	Finite volume mesh	147
8.2	Finite volume FD- SP_N approximation	149
8.2.1	Re-writing of the FD- SP_3 system	151
8.2.2	Finite volume derivation of FD- SP_3 system	152
8.2.3	Finite volume derivation of the boundary equations	154
8.2.4	Partial current operator	155
8.3	Algebraic formulation	156
8.3.1	Forward model	156
8.3.2	Partial current	161
8.3.3	Numerical solver for the forward model	162
8.4	Validation of FV-FD- SP_N algorithm	163
8.4.1	Numerical phantoms	163
8.4.2	Numerical study - phantom 1: homogeneous media	167
8.4.3	Numerical study - phantom 2: absorbing inclusions	172
8.4.4	Numerical study - phantom 3: scattering inclusions	178
8.4.5	Numerical study - phantom 4: absorbing and scattering inclusions	182
8.5	Discussion	186
9	Finite volume SP_N Model: Inverse Problem	188
9.1	Quasi-Newton methods	192
9.2	Unconstrained optimization	196
9.2.1	Gradient of objective function	197
9.2.2	Merit function	199
9.2.3	Unconstrained algorithm	199
9.3	PDE-Constrained optimization	201
9.3.1	Reduced Hessian approach	206
9.3.2	Summary of formulations	210

9.3.3	Merit function	211
9.3.4	rSQP PDE-constrained algorithm	212
9.4	Gradient of forward model	214
9.5	Validation and numerical simulations	217
9.5.1	Numerical phantoms	217
9.5.2	Performance evaluation	218
9.5.3	Disk phantom: case 1	219
9.5.4	Disk phantom: case 2	221
9.5.5	Disk phantom: case 3	223
9.5.6	Disk phantom: case 4	225
9.5.7	Disk phantom: case 5	226
9.5.8	Finger phantom	228
9.6	Discussion	233
10	Application Of The PDE-Constrained SP_N Algorithm To Clinical Data	235
10.1	Objective function for use with clinical data	236
10.2	Clinical data	239
10.3	Reconstruction results	241
10.4	Analysis of SP_N reconstruction data with CAD	245
10.5	Discussion	252
11	Conclusion	254
11.1	Diagnosis of RA with CAD techniques	254
11.2	Image reconstruction algorithm	258
11.3	Future work	261
11.4	Discussion	263
12	Bibliography	264
A	Forward Model	288
A.1	Conversion of FEM mesh into FVM mesh	288
B	Inverse Problem	293
B.1	L-BFGS algorithm	293
B.2	Right hand side of adjoint equation	294
B.3	Partial derivative of u	295
B.4	Gradient of objective function	296
B.5	Reduced gradient for rSQP method	297
B.6	Update direction for the forward variable	299
B.7	Merit function	300
B.8	Right hand side of adjoint equation: clinical data	300
C	Block-Structured Grids	302
C.1	Grid generation	304
C.2	Discretization of FD-ERT	306
C.2.1	Discretization on single grid	306

C.2.2	Discretization of FD-ERT on BSGs	307
C.3	Numerical phantoms	309
C.4	Results	311
C.5	Discussion	313

List of Tables

1.1	Definition of common acronyms.	7
2.1	Definition of variables.	13
2.2	Boundary and current coefficients for the SP_3 model as defined by Klose and Larsen in [1].	26
3.1	Definitions of the ACR 1987 revised criteria for classification of RA [2].	37
5.1	Six-group diagnostic table based on clinical evaluation and radiological imaging. .	58
5.2	Three-group diagnostic table based on clinical evaluation and radiological imaging.	72
5.3	Mean and standard deviation of individual features extracted from images of joints from groups S and H.	80
5.4	ROC analysis results using image features derived from μ_a and μ'_s images (data at 600 MHz).	81
5.5	LDA classification results using image features derived from μ_a and μ'_s distributions obtained with imaging data generated at a modulation frequency of 600 MHz. The numbers in bold indicate the largest values.	82
5.6	LDA classification results using image features derived from μ_a and μ'_s distributions obtained with imaging data generated at a modulation frequency of 300 MHz. The numbers in bold indicate the largest values.	84
5.7	LDA classification results using image features derived from μ_a and μ'_s distributions obtained with imaging data generated at a modulation frequency of 0 MHz. The numbers in bold indicated largest values observed.	84
6.1	Diagnostic table based on clinical evaluation and radiological imaging.	88
6.2	Summary of data pre-processing nomenclature.	92
6.3	Definition of basic statistical features.	94
6.4	Definition of GMM features. (* These feature are only applicable when considering three-dimensional images.)	95
6.5	Definition of FFT features.	97
6.6	Dunn's test for sample features, showing statistically significant differences between subgroups.	108

7.1	List of features with top Youden indices from ROC curve analysis (presented in Chapter 6).	117
7.2	ICC and ESS values of the features listed in Table 7.1.	135
7.3	Classification results with the KNN algorithm using 1, 5, and 11 nearest neighbors.	136
7.4	Classification results with the DA algorithm using linear and quadratic functions.	137
7.5	Classification results with the SOM algorithm using neurons $n = 9, 16$, respectively ($l = 1.0$).	137
7.6	Classification results with the SVM algorithm using linear, quadratic, and polynomial kernels.	138
8.1	Definition of variables.	146
8.2	Boundary and current coefficients for the SP_3 model as defined by Klose and Larsen in [1].	153
8.3	Definition of variables for the linear problem of the SP_3 model.	157
8.4	Background properties of the numerical phantoms. The modulation frequency is from the relation $\omega = 2\pi f$.	164
8.5	Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 1.	167
8.6	Computation time for phantom 1.	170
8.7	Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 2.	172
8.8	Computation time for phantom 2.	177
8.9	Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 3.	178
8.10	Computation time for phantom 3.	182
8.11	Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 4.	182
8.12	Computation time for phantom 4.	186
9.1	Definition of inverse problem variables.	191
9.2	Definition of PDE-constrained optimization terms.	203
9.3	Correlation coefficient between reconstruction and analytical solutions for the first phantom.	220
9.4	Correlation coefficient between reconstruction and analytical solutions for the second phantom.	222
9.5	Correlation coefficient between reconstruction and analytical solutions for the third phantom.	224
9.6	Correlation coefficient between reconstruction and analytical solutions for the fourth phantom.	226
9.7	Correlation coefficient between reconstruction and analytical solutions for the fifth phantom.	227
9.8	Correlation coefficient between reconstruction images and analytical solution for the first finger phantom (one absorbing object with $\mu_a = 0.05 \text{ cm}^{-1}$ and one scattering object with $\mu'_s = 10.0 \text{ cm}^{-1}$).	230
9.9	Correlation coefficient between reconstruction images and analytical solution for the second finger phantom (one absorbing object with $\mu_a = 0.25 \text{ cm}^{-1}$ and one scattering object with $\mu'_s = 7.5 \text{ cm}^{-1}$).	230
10.1	Summary of values assigned to reconstruction variables.	242

10.2	Classification results for SP_N and ERT based reconstructions with the k -fold method.	248
10.3	Name of optimal features selected during the training phase of the k -fold method using SP_N and ERT based reconstructions. Refer to Section 6.3.4 for a detailed explanation of the naming convention adopted for referencing image features. . . .	250
C.1	Properties of embedded fluorescent probes.	310
C.2	Summary of parameter values used for simulations.	310
C.3	Computation time, MPE, and RSU on a mouse cross-section.	312
C.4	Computation time, MPE, and RSU on a cylindrical phantom	314

List of Figures

1.1	(a) DOT imaging hardware. (b) Results from DOT reconstructions. (c) Computer aided diagnosis.	2
2.1	Absorption coefficient of oxygenated and deoxygenated hemoglobin (blood), water, and fat in the visible and NIR spectrum. Data for hemoglobin and fat was obtained from [3], citing [4,5] for hemoglobin data and [6] for fat data. Data for water was obtained from [7], citing [8] as the original source.	10
2.2	Illustration of a typical mammalian cell. Internal cell structures are labeled: 1) nucleolus, 2) nucleus, 3) ribosome, 4) vesicle, 5) rough endoplasmic reticulum, 6) Golgi apparatus, 7) cytoskeleton, 8) smooth endoplasmic reticulum, 9) mitochondrion, 10) vacuole, 11) cytosol, 12) lysosome, 13) centriole, 14) cell membrane. . .	11
2.3	(a) Example of laser light illuminating a human finger with NIR photons. (b) A typical distribution of exiting photons at the surface of the finger (i.e. a “red glow”).	12
3.1	(a) Example of symmetric synovitis affecting the hands, highlighting the location of a PIP joint. (b) X-ray images with radiographic evidence of RA, highlighting the location of a PIP joint. (c) Illustration of a typical healthy joint. (d) Illustration of a typical joint exhibiting symptoms of RA. (c,d) Schematic of a typical synovial joint. These joints are surrounded by a “joint capsule” that provides protection, the inside of which is called the synovial membrane or synovium. The synovial capsule is filled with synovial fluid, a normally clear and colorless viscous non-Newtonian fluid whose primary purpose is to provide lubrication to the articulating surface of the joint. The onset of symptoms from RA causes the synovium to become inflamed, causing warmth, redness, swelling, and pain. Progression of the disease can lead to invasion of the inflamed synovium and result in damage to the articulating surfaces and surrounding bone. The synovial fluid undergoes substantial changes, resulting in yellow to red color and cloudy to opaque clarity. Image courtesy of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (http://images.niams.nih.gov/detail.cfm?id=74 - last accessed June 13, 2012).	35
5.1	Bones of the human hand, highlighting (in red) the PIP joints (II-IV) imaged with FD-DOT. For reference, PIP joint I is the thumb.	55

5.2	(a-b) The two figures show T1-weighted images before (a) and after (b) administration of a bolus of Gd-DTPA. The regions marked with a white circle shows synovitis in a PIP joint. (c-d) The two figures show T1-weighted images before (c) and after (d) administration of a bolus of Gd-DTPA. The region marked with a white circle shows erosions in all three joints inside the circle.	56
5.3	Ultrasound images of a healthy joint (a), and joints affected by RA (b,c,d). The larger the anechoic area and/or extent of synovial hypertrophy, the higher was USS (USS = 1 minimal, USS = 2 moderate, USS = 3 extensive effusion/hypertrophy). Images were produced by placing a hockey stick linear array transducer on the palmar side of the PIP joint. (E \equiv effusion, S \equiv synovitis, T \equiv tendon, JC \equiv joint cavity)	58
5.4	Set-up of tomographic scanning unit, schematic (a) and photograph (b). The hardware parts shown are the (1) laser diode, (2) laser diode driver, (3,7) signal generator, (4) finger, (5) focusing lens, (6) ICCD camera, (8) high rate imager, and (9) computer.	59
5.5	(a) Surface registration and (b) three-dimensional mesh generation. The surface scanning unit detects the shape of the laser line on the finger surface and determines the three-dimensional-surface coordinates while the laser line is scanned over the finger. The background serves for calibration of the camera coordinate system. Surface registration and three-dimensional mesh generation; (a) three-dimensional laser scanning to obtain a finger joint geometry and (b) three-dimensional finite volume mesh generated using the laser scanned surface mesh.	60
5.6	Position of 11 sources (a) and detectors (b) on the posterior (dorsal) and anterior (palmar) surface of a sample PIP joint.	62
5.7	Transillumination captured by the ICCD detector unit on the posterior (palmar) surface of a PIP joint from a subject with RA (a) and a healthy control (b).	62
5.8	(a) Example of how TP, TN, FP, and FN can be computed given a threshold value (vertical line). Se and Sp can be computed from those values. (b) By continuously varying the threshold value, Se and Sp pairs can be computed and used to plot ROC curves. The 45° line is shown for reference.	70
5.9	(a) Photo showing the approximate size of a finger for which an FVM mesh is generated and used to reconstruct the FD-DOT image. (b) Example of a two-dimensional cross-section through a three-dimensional reconstruction of the absorption coefficients in a finger joint. Also shown is the ROI from which various optical parameters are determined. The ROI is limited to regions at least 2 mm away from boundaries and within the lateral extent of the source and detector location. (c) For further orientation of the imaging range, the corresponding sagittal cross-section from an MRI scan is shown.	71
5.10	Schematic of the feature extraction process. Four features are extracted from the absorption and scattering images of each joint, resulting in a vector of eight image features.	72
5.11	Examples of two-dimensional FD-DOT cross-sections through (a) healthy joints and (b) joints affected by RA, showing the spatial distribution of μ_a . Imaging was performed with a source modulation frequency of 600 MHz.	75

5.12	Examples of two-dimensional FD-DOT cross-sections through (a) healthy joints and (b) joints affected by RA, showing the spatial distribution of μ'_s . Imaging was performed with a source modulation frequency of 600 MHz.	75
5.13	Mean values with 95% confidence intervals of all features (600 MHz data) from all six diagnostic subgroups (A, B, C, D, E, and H in Table 5.1). The left column corresponds to features from μ_a images, while the right column corresponds to features from μ'_s images. If there is no overlap between groups, then the mean of the given feature is different from the other means at statistically significant levels ($p < 0.05$).	77
5.14	Mean values with 95% confidence intervals of all features extracted from FD-DOT images (600 MHz) from groups S, A, and H (see Table 5.2). The left column corresponds to features from μ_a images, while the right column corresponds to features from μ'_s images. If there is no overlap between groups, then the mean of the given feature is different from the other means at statistically significant levels ($p < 0.05$).	78
5.15	Distribution of features found in fingers of healthy (blue dots, group H), symptomatic (red dots, group S), and asymptomatic (gray dots, group A) joints. The bars show the standard error with respect to $var(\mu_a)$ and $max(\mu_a)$, respectively, for each group.	79
5.16	ROC curves for the features max , min , var , and $ratio$ extracted from μ_a and μ'_s images obtained at 600 MHz, respectively.	81
5.17	Example of LDA decision boundary for discriminating between groups S (red) and H (blue) based on features $var(\mu_a)$ and $max(\mu_a)$	83
6.1	Visualization of typical finger geometries. Sample μ'_s cross-sections are visualized within the finger geometry: (a) coronal (xy), (b) sagittal (xz), and (c) transverse (yz) planes.	89
6.2	(a-b) Absorption and (c-d) scattering coronal cross-sections of PIP joints from subjects without (a,c) and with RA (b,d). All images are reconstructed from data obtained with 600 MHz source modulation.	90
6.3	An example of the summation of coronal, transverse, and sagittal slices of the three-dimensional data set to create new data sets (a) SC , (b) SS , and (c) ST	91
6.4	Data processing steps, starting with the raw CCD data, followed by processing of the unstructured and structured data sets, and ending with application of the 3 projection operators to the three-dimensional structured data. This results in 11 distinct data sets (represented by each circle and summarized in Table 6.2). The feature extraction operators are applied to each set.	93
6.5	(a-b) Three-dimensional example of μ_a around a joint. (c-d) Coronal cross-section of μ_a across the same PIP joint. (a) Iso-surfaces of the original three-dimensional data. (b) Iso-surfaces of the GMM model showing a good approximation to the original data. (c) Isolines are superimposed on the original image to show the resulting fit from GMMs. (d) The model image generated from the coefficients of the GMM model.	95

6.6	FFT coefficients of a two-dimensional image A . Unique coefficients are numbered in increasing order according to distance from the origin, with $z > y > x$ axis preference used as a tie breaker.	97
6.7	(Top row) Sagittal, coronal, and transverse cross sections of a healthy PIP joint. (Middle row) GMM parameterization approximation. (Bottom row) Reconstruction of the original image using only the extracted FFT coefficients of the first five frequencies.	103
6.8	H -statistic value for μ_a (top two rows) and μ_s (bottom two rows) features. The threshold values for H at the 0.05 and 0.001 confidence levels are 11.07 and 20.52, respectively, and shown as horizontal red lines.	105
6.9	Mean value and standard error of the maximum, minimum, mean, and variance of μ_a and μ'_s images. A two-sample <i>student-t</i> test shows differences between features from subjects with RA and without RA are statistically significant at the $\alpha = 0.01$ level. The variance is scaled to display on the same axis (\dagger scaled by 100; \ddagger scaled by 10).	107
6.10	(a-b) ROC curves for basic features from unstructured data (F01:UV:a and F01:UV:s, F02:UV:a and F02:UV:s, etc.), (c) absolute error between original data and GMM model features, and (d) DC component of two-dimensional-FFT features. FFT features perform better, in terms of area under the curve, than basic features and GMM coefficients.	109
6.11	Single feature classification results reported as the Youden index ($Se + Sp - 1$) from ROC analysis for μ_a (top two rows) and μ'_s (bottom two rows) features.	111
7.1	Distribution of sample features. In (a) the six distinct diagnosis groups (A, B, C, D, E, H) are identified, while in (b) the five cohorts diagnosed with RA (A, B, C, D, E) are grouped into one “RA” group.	118
7.2	(a) KNN, (b) QDA, (c) LDA, and (d) SVM boundaries for a non-linearly separable two-class problem. (a) Examples of $k = 1$ and $k = 13$ for KNN are shown for a new data point p , denoted by \blacktriangle . (d) Support vector data points from classes 1 and 2 are denoted by white square or circle markers, respectively.	119
7.3	(a) Example of linearly separable two-class problem that can be separated by infinite different lines, but there is only one optimal separating line. The “margin” corresponding to each acceptable solution (separating line) is denoted with dashed lines. In this example, solution 2 is preferred over solution 1 because of the size of the margin. (b) Example of non-linearly separable two-class problem. The “slack” variable ξ is highlighted.	123
7.4	(a) Sample within-generation values of the augmented Youden index Y^* for all possible feature combinations (mutants). (b) Sample evolution of objective function over multiple generations.	132
7.5	(a) LDA, (b) QDA, and (c) SVM decision boundaries separating affected from healthy data using image features 5 and 19. Support vector data points are identified by dots with white filling.	133
7.6	Sample Se and Sp evolution paths obtained with KNN, DA, SOM, and SVM classifiers. The lower bounds (Lb) for the 95% confidence interval are shown for Se and Sp , respectively. The upper bound in these cases is 100.0%.	134

7.7	Frequency with which all features appear as optimal classifiers using KNN, DA, SOM, and SVM.	139
8.1	(a) An example physical object \mathbf{O} , (b) and its corresponding discretization into triangular elements (red) and finite volume (FV) elements (blue). Each surface of the FV element has a corresponding normal vector $\hat{\mathbf{n}}$	148
8.2	Geometric center (blue dot) computed with the (a) median and (b) Delaunay methods.	149
8.3	Discretization of a disk (2.0 cm diameter) into (a) triangular elements (i.e. FEM mesh) and (b) into finite-volume elements (i.e. FVM mesh). (c) A single boundary source (red dot) and 50 detectors (blue dots) are distributed around the boundary of the phantom. The internal cross-section of interest consists of mesh points 0.5 ± 0.01 cm from the center of the phantom, represented by the inner circle.	163
8.4	(a-d) Definition of two-dimensional phantoms number 1 through 4, respectively. (a) Homogeneous and (b-c) inhomogeneous phantoms. All simulations on these phantoms are performed with anisotropic factor $g = 0.95$, source modulation frequency $\omega = 2\pi f = 600$ MHz, and refractive index $n_m = 1.0$	165
8.5	[Phantom 1] Amplitude (log scale) of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	168
8.6	[Phantom 1] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	169
8.7	[Phantom 1] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2	169
8.8	[Phantom 1] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.	170
8.9	[Phantom 1] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.	171
8.10	[Phantom 2] Amplitude of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	173
8.11	[Phantom 2] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	174
8.12	[Phantom 2] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2	175
8.13	[Phantom 2] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.	176

8.14	[Phantom 2] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.	177
8.15	[Phantom 3] Amplitude of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	179
8.16	[Phantom 3] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	179
8.17	[Phantom 3] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2	180
8.18	[Phantom 3] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.	181
8.19	[Phantom 3] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.	181
8.20	[Phantom 4] Amplitude of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	183
8.21	[Phantom 4] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).	184
8.22	[Phantom 4] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2	184
8.23	[Phantom 4] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.	185
8.24	[Phantom 4] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.	185
9.1	(a) Location of boundary sources (red). (b) Location of boundary detectors (blue) for a single boundary source (red) and placement of the internal cross-section of interest (black).	218
9.2	(Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).	219
9.3	Values of (a) absorption and (b) scattering coefficient along the internal cross section defined in Fig. 9.1b.	220

9.4	Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.	221
9.5	(Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).	222
9.6	Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.	223
9.7	(Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).	224
9.8	Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.	225
9.9	(Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).	225
9.10	Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.	226
9.11	(Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).	227
9.12	Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.	228
9.13	(a) Position of the 11 sources and (b) location of the 155 detectors on the posterior and anterior surface of the finger phantom, respectively.	229
9.14	Definition of finger-like phantoms; both phantoms have one absorbing (red object, top row) and one scattering (blue object, bottom row) inclusion in a homogeneous background. The optical properties of the first phantom are as follows; the background properties are $\mu_a = 0.01 \text{ cm}^{-1}$ and $\mu'_s = 12.5 \text{ cm}^{-1}$ ($g = 0.95$), the absorbing inclusion is $\mu_a = 0.05 \text{ cm}^{-1}$, and the scattering region is $\mu'_s = 12.5 \text{ cm}^{-1}$. The second finger phantom is similar, except for differences in the optical properties of the background ($\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$, $g = 0.95$), the absorbing inclusion ($\mu_a = 0.25 \text{ cm}^{-1}$), and the scattering region ($\mu_s = 7.5 \text{ cm}^{-1}$).	229

9.15	Three-dimensional reconstruction images corresponding to the second finger phantom. Depicted are the absorption inclusion (red object) and scattering void (blue object) in otherwise homogeneous media. The true optical properties of the absorbing inclusion is $\mu_a = 0.25 \text{ cm}^{-1}$, the scattering region is $\mu_s = 7.5 \text{ cm}^{-1}$, and the background are $\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$. The reconstruction algorithm is able to accurately locate the inhomogeneities and estimate their optical properties. The phantom is defined in Fig. 9.14.	231
9.16	Cross-sectional slices inside the finger phantom of the reconstruction images of the absorption (top row) and scattering (bottom row) coefficients obtained with the PDE-constrained SP_3 models for the second finger phantom. The true optical values of this phantom are as follows; background media is defined with $\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$ ($g = 0.95$), the absorption inclusion has $\mu_a = 0.25 \text{ cm}^{-1}$, and the scattering inclusion has $\mu_s = 7.5 \text{ cm}^{-1}$	232
9.17	Performance profile of reconstruction algorithms. Objective function value over (a,c) time and (b,d) at the conclusion of each inverse iteration for the (top row) first and (bottom row) second finger phantom.	232
10.1	An illustration of a typical finger section. The PIP joint is located approximately in the middle of the object, whose length is approximately 4 cm. (a) The position of the 11 unique sources are indicated on the posterior surface of the finger. (b) The position of the detectors (i.e. mesh nodes that map directly to a pixel on the CCD camera-based detection system) are shown on the anterior surface of the finger section.	240
10.2	Transillumination captured by the ICCD detector unit on the posterior (palmar) surface of a PIP joint belonging to (a) a subject with RA and (b) a healthy control. .	240
10.3	(a-b) Absorption and (c-d) scattering cross-sections of distinct joints from subjects with and without RA.	243
10.4	(a-b) Absorption and (c-d) scattering cross-sections of distinct joints from subjects with and without RA.	244
10.5	Flow chart of k -fold cross-validation.	246
10.6	Flow chart of the data mining process that selects optimal features and the best classification algorithm for diagnosing RA from DOT images.	247
10.7	Mean value and standard error of the optimal features selected from analysis of the (a) ERT and (b) SP_3 based reconstruction images. A two-sample <i>student-t</i> test shows that differences between features from subjects with RA and without RA are statistically significant at the $\alpha = 0.01$ level. (a) Features denoted by \dagger are scaled by a factor of 10. (b) Features denoted by \dagger and \ddagger are scaled by a factor of 50 and 500, respectively.	250
C.1	General sequence in which major subroutines are executed in our algorithm. . . .	304
C.2	Visualization of algorithm for determining computational domains.	305

C.3	Examples of the increased boundary resolution attained with BSGs relative to a single coarse grid (b,c,d) and reduction in mesh density attained by a BSG relative to a fine grid (d,e). The geometry the grids approximate is the cross-section of a mouse obtained from a segmented MRI image (a). Depicted in (b-d) are 1-, 2-, and 3-level BSGs, while (e) is the fine grid discretization of the mouse cross-section.	305
C.4	(a) Example of possible grid points on interior boundaries. There are five distinct cases. The first four cases are labeled i-iv and the fifth case is specified by the black dot. (b) The grid point denoted by the black diamond is an interior boundary point; the fine grid is to its left and the coarse grid is to its right. The black dot is the virtual point that must be created when the differencing scheme in the x-direction requires a forward Euler-step. Solutions at the four white dots surrounding the black dot are averaged to create the virtual point.	308
C.5	(a) disk, (b) 2D mouse cross section, (c) cylinder, and (d) 3D mouse model. The positions of boundary sources are shown for as black dots, while detectors are assigned to every boundary point. (d) Sources for the mouse are located within the interval defined by $x = [16, 18]$, $y = [43, 46]$, and $z = 0$. Similarly, detectors are defined within $x = [16, 18]$, $y = [43, 46]$, and $z = 15$	309
C.6	(a-c) Examples of single structured grids fitted to the disk, cylinder, and mouse phantom, respectively. (d-f) 2-level BSGs fitted to same three phantoms. Three-dimensional BSGs and their interior structure are visualized by showing only a section of the full three-dimensional shape (e,f).	311
C.7	Three-dimensional representations of excitation (a) and emission (b) fluence on the cylindrical phantom. (d) Sample excitation on the three-dimension mouse phantom.	313

Publications Resulting From This Work

PATENTS

1. A.H. Hielscher H, H.K. Kim, **L.D. Montejo**, “Medical imaging devices, methods, and systems,” U.S. Patent Application No. 13/993,592, filed June 12, 2013.
2. A.H. Hielscher, **L.D. Montejo**, H.K. Kim, “Combined PDE-constrained and simplified spherical harmonics algorithm for image reconstruction,” U.S. Provisional Application No. 61/729,255, filed November 21, 2012.
3. A.H. Hielscher H, H.K. Kim, **L.D. Montejo**, S. Blaschke, U.J. Netz, G.A. Mueller, J. Beuthan, “Medical Imaging Devices, Methods, and Systems,” Jun, 21 2012: WO/2012/082804 (PCT/US2011/064723).

JOURNAL PAPERS

1. **L.D. Montejo**, Jia, H.K. Kim, A.H. Hielscher, Frequency-domain DOT image reconstruction algorithm with the SP3 light propagation model,” JQSRT, under-review (Dec, 2013).

2. **L.D. Montejo**, J. Jia, H.K. Kim, U.J. Netz, S. Blaschke, G.A. Muller, A.H. Hielscher, "Computer-aided diagnosis of rheumatoid arthritis with optical tomography, Part 1: feature extraction," *J. Biomed. Opt.* 18 (7), 076001 (July 15, 2013). [doi: 10.1117/1.JBO.18.7.076001]
3. **L.D. Montejo**, J. Jia, H.K. Kim, U.J. Netz, S. Blaschke, G.A. Muller, A.H. Hielscher, "Computer-aided diagnosis of rheumatoid arthritis with optical tomography, Part 2: image classification," *J. Biomed. Opt.* 18 (7), 076002 (July 15, 2013). [doi: 10.1117/1.JBO.18.7.076002]
4. Hielscher AH, Kim HK, **L.D. Montejo**, Blaschke S, Netz UJ, Zwaka PA, Illing G, Muller GA, Beuthan J, "Frequency-domain optical tomographic imaging of arthritic finger joints," *IEEE Trans Med Imaging*. 2011 Oct;30(10):1725-36.
5. **L.D. Montejo**, A.D. Klose, A.H. Hielscher, "Implementation of the equation of radiative transfer on block-structured grids for modeling light propagation in tissue," *Biomedical Optics Express*, Vol. 1, Issue 3, pp. 861-878 (2010). (doi:10.1364/BOE.1.000861)

CONFERENCE PROCEEDINGS

1. **L.D. Montejo**, J. Jia, H.K. Kim, A.H. Hielscher, Evaluation of Fourier transform coefficients for the diagnosis of rheumatoid arthritis from diffuse optical tomography images," *Proc. SPIE*. 8578, Optical Tomography and Spectroscopy of Tissue X 85781N (March 25, 2013). [PDF, Link, doi: 10.1117/12.2004846]
2. J. Jia, J. Lee, **L.D. Montejo**, H.K. Kim, A.H. Hielscher, Measurement operator for angular dependent photon propagation in contact-free optical tomography," *Proc. SPIE* 8578, Optical Tomography and Spectroscopy of Tissue X, 857815 (March 25, 2013). [PDF, Link, doi:10.1117/12.2003200]
3. H.K. Kim, M.A. Khahil, J. Gunther, **L.D. Montejo**, A.H. Hielscher, A dynamic image reconstruction method with spatio-temporal constraints," *Proc. SPIE* 8581, Photons Plus Ultrasound: Imaging and Sensing 2013, 858131 (March 4, 2013). [PDF, Link, doi:10.1117/12.2003394]
4. **L.D. Montejo**, J. Jia, H.K. Kim, A.H. Hielscher, Classification of OT Images of Arthritic Joints Using Spatial-Fourier Frequency Coefficients," in *Biomedical Optics and 3-D Imaging*, OSA Technical Digest (Optical Society of America, 2012), paper BTu3A.38.
5. **L.D. Montejo**, H.K. Kim, A.H. Hielscher, A finite-volume algorithm for modeling light transport with the time-independent simplified spherical harmonics approximation to the equation of radiative transfer," *Proc. SPIE* 7896, Optical Tomography and Spectroscopy of Tissue IX, 78960J (February 17, 2011). [PDF, Link, doi:10.1117/12.875967]
6. **L.D. Montejo**, H.K. Kim, Y. H€lme, J. Jia, J.D. Montejo, U.J. Netz, S. Blaschke, P.A. Zwaka, G.A. Meller, J. Beuthan, A.H. Hielscher, Analysis and classification of optical tomographic images of rheumatoid fingers with ANOVA and discriminate analysis," *Proc. SPIE* 7890, Advanced Biomedical and Clinical Diagnostic Systems IX, 78900L (February 21, 2011). [PDF, Link, doi:10.1117/12.875977]

7. **L.D. Montejo**, A.D. Klose, A.H. Hielscher, “Modeling fluorescence light propagation in arbitrarily shaped domains with the equation of radiative transfer on block-structured grids,” in Biomedical Optics, OSA Technical Digest (CD) (Optical Society of America, 2010), paper BWF2.
8. **L.D. Montejo**, J.D. Montejo, H.K. Kim, U.J. Netz, C.D. Klose, S. Blaschke, P.A. Zwaka, G.A. Miller, J. Beuthan, A.H. Hielscher, “Comparison of Classification Methods for the Detection of Rheumatoid Arthritis with Optical Tomography,” in Biomedical Optics, OSA Technical Digest (CD) (Optical Society of America, 2010), paper BSuD38.
9. A.H. Hielscher, H.K. Kim, U. Netz, **L.D. Montejo**, C.D. Klose, S. Blaschke, P.A. Zwaka, G.A. Miller, J. Beuthan, “Frequency-Domain Optical Tomography of Arthritic Joints,” in Biomedical Optics, OSA Technical Digest (CD) (Optical Society of America, 2010), paper BSuD87.
10. **L.D. Montejo**, A.D. Klose, A.H. Hielscher, “Algorithm for the Frequency Domain Radiative Transfer Equation on Block-Structured Grids,” from the 36th Annual Northeast Bioengineering Conference, available from IEEE, 2010.
11. **L.D. Montejo**, J.D. Montejo, H.K. Kim, U.J. Netz, C.D. Klose, S. Blaschke, P.A. Zwaka, G.A. Miller, J. Beuthan, A.H. Hielscher, “Diagnosis of Rheumatoid Arthritis with Optical Tomography: Comparison of Classification Methods,” from the 36th Annual Northeast Bioengineering Conference, available from IEEE, 2010.
12. **L.D. Montejo**, A.D. Klose, A.H. Hielscher, “Implementation of the radiative transfer equation on block-structured grids for modeling fluorescence light propagation in tissue with arbitrary shape,” Proc. SPIE, Vol. 7174, 717410 (2009).
13. **L.D. Montejo**, A.D. Klose, Numerical and Experimental Studies of GFP Expressing Small Animal Models with the Equation of Radiative Transfer, World Molecular Imaging Congress, Optics III, presentation 1073 (2009).

Acknowledgements

I have been very fortunate to have had great role models, mentors, and friends throughout my academic career and I am extremely grateful to all. I would first like to thank my dissertation committee for all the help and the advice that has allowed me to complete this dissertation; Professors Andrew F. Laine, Andreas H. Hielscher, Paul Sajda, Elizabeth M.C. Hillman, and Hyun K. Kim.

I would like to specifically thank my doctoral advisor, Professor Andreas H. Hielscher, for giving me the opportunity to pursue graduate studies and for the continuous support throughout these past six years. I cannot imagine being able to complete this dissertation without his calming influence and constant support throughout the good times, and especially the bad times. I am forever grateful to my close collaborators, Professors Hyun K. Kim and Alexander D. Klose, for their patience and willingness to work with me during the early stages of my career. I am sure that I was more of a distraction than help at that point!

I sincerely thank all current and past members of the Biophotonics and Optical Radiology Laboratory for their contributions to this work and their friendship throughout the years; James Masciotti, Xuejun Gu, Molly Flexman, Jong-Hawn Lee, Michael Khalil, Jingfei Jia, Chistopher Fong, Jacqueline Gunther, and Jennifer Hoi. I am grateful to Jingfei Jia, Yrjö Häme, and Julio D. Montejo for contributing to this work at various stages. I would also like to express my gratitude to our clinical collaborators for providing the clinical data used in this dissertation; Sabine Blaschke (Dept. of Nephrology and Rheumatology, University Medicine of Göttingen, Germany), Uwe J. Netz (Dept. of Medical Physics and Laser-Medicine, Charité, University Medicine Berlin and Laser-und Medizin-Technologie GmbH, Germany), Paul A. Zwaka (Dept. of Radiology, Georg-August Medical University, Germany), Gerd Illing (Laser-und Medizin-Technologie GmbH, Germany), Gerhard A. Müller (Dept. of Nephrology and Rheumatology, University Medicine of Göttingen, Germany), and Jürgen Beuthan (Dept. of Medical Physics and Laser-Medicine, Charité, University Medicine Berlin, Germany).

I am grateful to all teachers and professors that have contributed to my education over the past 25 years. My decision to pursue graduate studies was primarily motivated by the experience of

spending the summer of 2005 as an undergraduate research assistant at the Mathematical and Theoretical Biology Institute (MTBI) at the Los Alamos National Laboratory (LANL). I am eternally grateful to Professor Carlos Castillo Chavez (Arizona State University) for the opportunity to attend MTBI, for introducing me to the world of academic research, and for his continual guidance. I am also tankful to all my colleagues and mentors that helped make summer research at MTBI invaluable and unforgettable. Professor David Keyes (Dean at King Abdullah University of Science and Technology) was crucial in my decision to pursue graduate studies at Columbia University and an invaluable mentor through my first years of research while he was a Professor of Applied Mathematics at Columbia University.

Lastly, I am eternally grateful to my fiancé, Lily, for the uncountable ways in which she has been crucial to my life and my success in graduate school.

Dedication

To my family, with love.

Introduction

Decades of research into the development of Diffuse Optical Tomography (DOT) imaging have been with one aim: to establish DOT as tool for aiding in the monitoring and the detection of physiological processes and pathologies. The goal of this dissertation is the development and implementation of algorithm-based solutions for improving the clinical viability and research utility of DOT, with a particular focus on its application to the diagnosis Rheumatoid Arthritis (RA).

DOT is a novel medical imaging modality where near-infrared (NIR) photons are used to probe tissue. NIR photons are injected into the tissue of interest, typically by focusing a non-invasive NIR laser onto the surface, where the photons interact with the underlying physiology. These interactions result in absorption or scattering events, where photons are either annihilated or redirected from their incoming direction, respectively. Scattered photons may eventually escape the tissue of interest. Information gathered from these exiting photons can be used to reconstruct the spatial distribution of physiologically relevant parameters such as tissue absorption and scattering coefficients, concentration of oxygenated and deoxygenated hemoglobin, and concentration of fluorescent and bioluminescent probes.

Over the last two decades, the field of DOT has made significant developments in imaging hardware [9, 10, 11, 12, 13, 14, 15] and image reconstruction software [1, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. DOT has progressed from purely theoretical studies and bench-top experiments, to clinical trials that explore its utility in a wide variety of therapeutic areas. Traditionally, DOT has

been primarily applied to imaging of the breast [11, 15, 28, 29] and brain [30, 31, 32, 33, 34]. More recently, studies have also focused on imaging of finger joints [9, 10, 12, 35].

A promising application of this new technology is to the diagnosis of RA through imaging of finger joints [10, 12, 13, 14, 36, 37]. The motivation for imaging finger joints with light is apparent. Unlike the breast or brain, fingers are relatively small and the transmitted light intensities are comparatively high, allowing for simple signal detection. Additionally, studies in the mid nineteen-nineties showed that the scattering properties of the synovial fluid inside joints changed as a result of the onset of RA, indicating that there is potential for natural optical contrast between healthy joints and joints with RA.

There are three general areas that characterize the state of DOT: imaging hardware, reconstruction algorithms, and image analysis tools (Fig. 1.1). The focus of this dissertation is on image reconstruction algorithms and image analysis tools for improved disease detection. This will be achieved by developing a general framework for incorporating DOT imaging into the diagnosis of RA and by making improvements to the image reconstruction process. This work will make DOT imaging more immediately useful to physicians, researchers, and practitioners.

While innovative imaging systems have been developed that measure light transmission signals with high accuracy, two significant problems persist in our ability to routinely use DOT as a tool for diagnosing RA: (1) there is no established framework for accurately diagnosing the presence

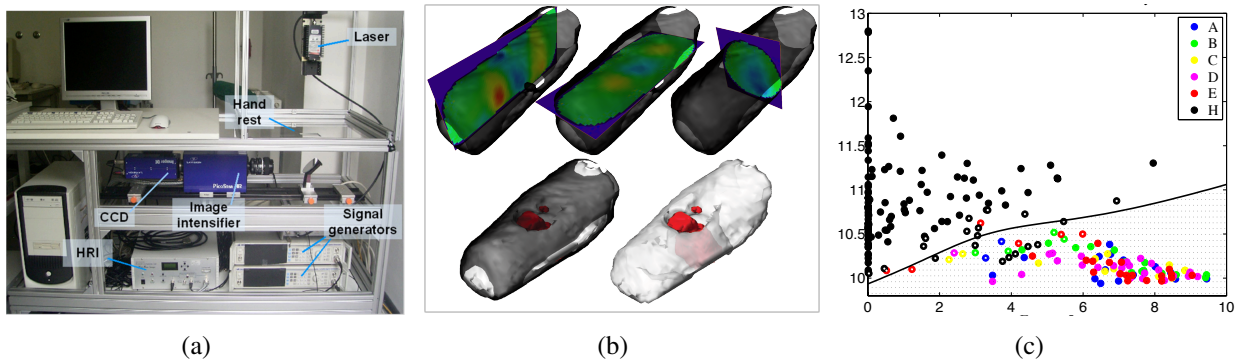


Figure 1.1: (a) DOT imaging hardware. (b) Results from DOT reconstructions. (c) Computer aided diagnosis.

of diseased tissue from DOT images, and (2) reconstruction of tissue properties and embedded inhomogeneities remain computationally demanding.

First, because DOT is such a young imaging modality and clinical results are still limited, there is, as of yet, no set standard for evaluating a DOT image for the presence of RA. As a result, reported sensitivities and specificities have not yet reached clinically acceptable levels. This lack of clinical evidence prevents DOT from gaining more mainstream acceptance among the medical community.

Secondly, while data acquisition is often fast (typically less than 15 minutes) and subjects discomfort low, the computational time required to obtain the optical properties of finger joints remains considerably long (> 3 hrs. per finger)¹ and requires significant computational resources (> 6 GB RAM per finger)², severely limiting the practical use of this modality. The main reason for the long computation times is the need to use the equation of radiative transfer (ERT) when modeling light propagation in finger joints. Traditionally, the ERT has been the preferred model to use when computing the optical properties of finger joints as it is known to be the most accurate deterministic model of light propagation in small volumes. ERT-based reconstruction algorithms, however, are highly complex and require substantial computational recourses.

This dissertation centers on the idea that a computer-aided diagnosis (CAD) toolbox and computationally efficient reconstruction algorithms can help transform DOT imaging into an important tool in the diagnosis of RA. An important underlying hypothesis is that RA affects the optical properties of the synovium and surrounding tissue to a degree that can be captured by DOT imaging. Equally importantly, I postulate that approximations to the ERT model, which can be substantially less expensive to solve, can be used in an image reconstruction algorithm to resolve the optical properties of finger joints with enough accuracy so as to allow for the accurate diagnosis of RA.

Thus, the overall goals of this dissertation are to show that RA does indeed create sufficient contrast in DOT images to accurately detect the presence of the disease and to develop algorithms

¹Computation time can be much longer than 3 hours, easily reaching over 12 hours per finger when numerically accurate light propagation models are used.

²As in the case of computation time, physical memory requirement can easily increase and is dependent on many aspects related to the computational domain and the light propagation model.

that will lead to the clinical acceptance of DOT for the diagnosis of RA. To achieve this, the following specific aims are proposed.

Aim 1: Develop a CAD toolbox for DOT that leads to significantly increased sensitivities and specificities in the diagnosis of RA. The toolbox works in two steps. First, heuristic features are extracted from DOT images. These features are used to classify images of finger joints as “affected” or “not affected” by RA. Classification is performed using traditional methods; such as linear discriminate analysis and ANOVA, as well as more recently developed machine learning approaches, such as support vector machines (SVM) and self organizing maps (SOM). The combination of features and classification algorithms that lead to the highest sensitivity and specificity in the diagnosis of RA is determined.

Aim 2: Develop algorithms that increase the computational efficiency and speed of the image reconstruction process used in DOT imaging of finger joints. This is achieved by developing a reconstruction algorithm based on the simplified spherical harmonics (SP_N) model for amplitude-modulated light propagation. The algorithm is based on a PDE-constrained, reduced Hessian, quadratic programming approach that leads to improved reconstruction speeds and memory requirements compared to standard DOT reconstruction techniques. We report results comparing the computational efficiency of the SP_N algorithm and the benchmark transport-based algorithm. In addition, an algorithm for solving the equation of radiative transfer (ERT) on block-structured grids (BSG) is developed to decrease the computational demand associated with using the transport model in DOT.

Aim 3: Validate the FV- SP_N reconstruction algorithm with clinical data. Existing clinical data of imaged finger joints is used to assess the performance of the SP_N reconstruction algorithm in a typical clinical setting. The best classification algorithm from Aim 1 is used to diagnose each SP_N -based joint reconstruction as affected or not affected with RA using the CAD toolbox

developed in Aim 1. Classification results obtained with the SP_N -based images are compared to diagnostic results from Aim 1.

The central thesis of this dissertation is that RA can be diagnosed with DOT and that CAD tools are necessary to enhance the clinical value of DOT images. Subsequent prospective clinical studies are needed to further validate these claims, however, this work provides the foundation necessary to show that DOT has the potential to become an important tool in identifying subjects with RA. Furthermore, the evidence presented in this work suggests that it may be possible for DOT to identify subjects with early stages of RA who might otherwise not qualify for a full evaluation under the guidelines of the 2010 criteria for classification of RA [38].

DOT imaging, if proven to be useful in diagnosing RA, could become an easy to use tool for physicians and other specialists. Evidence of the presence of RA in PIP joints from DOT images could be sufficient to warrant referral of a subject to a rheumatologist for evaluation and diagnosis. This would help fulfill a current need in the medical community, as identified by ACR/EULAR, which seeks a more robust method for identifying high-risk populations [38]. Furthermore, beyond allowing for the diagnosis of RA in well-established cases, it is possible that DOT can aid in the diagnosis of RA in subjects with early symptoms, as it is conceivable that changes in the optical properties of the peripheral joints start prior to manifestation of the symptoms necessary for diagnosis of RA with the current diagnostic criteria. This may be possible as the criteria, known as the ACR criteria, on its own, are not sensitive to the early onset of RA as subjects typically do not meet at least 4 of the 7 criteria necessary for a positive diagnosis [39].

If the methods proposed in this dissertation can be proven to work well with RA, then it is foreseeable that these methods can be expanded to include the diagnosis of several other types of inflammatory arthritis that are also known to alter the optical properties of the synovial fluid, including lupus arthritis and gout. Additionally, this technique may be expanded to detect atherosclerosis, which has been shown to be an inflammatory autoimmune disorder [40]. Detection of non-inflammatory arthritis, including osteoarthritis and arthritis of thyroid disease, might also be possi-

ble. Furthermore, the methods and techniques presented in this dissertation could also be deployed in the diagnosis of other diseases with DOT, including breast cancer and peripheral arterial disease (PAD).

The remainder of this dissertation is organized to provide a succinct introduction to DOT (Chapter 2) and rheumatoid arthritis (Chapter 3). Then, Chapters 5, 6, and 7 focus on work dedicated to the development of CAD techniques for the diagnosis of RA from DOT images. Details of the reconstruction algorithm based on the SP_N light propagation model are presented in Chapters 8 and 9. Results from application of the SP_N algorithm to clinical data are presented in Chapter 10. Finally, this dissertation concludes with Chapter 11, where a brief summery is followed by a discussion of future work and the impact of the work presented herein.

Table 1.1: Definition of common acronyms.

Acronym	Definition
ACR	American College of Rheumatology
BSG	Bock-structured grid
CE	Clinical evaluation
CRP	C-reactive protein
DA	Discriminate analysis
DAS	Disease activity score
DE	Diffusion equation
DMARD	Disease modifying anti-rheumatic drugs
DOT	Diffuse optical tomography
ERT	Equation of radiative transfer
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
FD-DOT	Frequency-domain diffuse optical tomography
FDM	Finite-differences method
FEM	Finite-element method
FFT	Fast Fourier transform
FN	False negative
FP	False positive
FVM	Finite-volume method
GMM	Gaussian mixture model
KNN	k nearest neighbors
LDA	Linear discriminate analysis
MC	Monte Carlo
MRI	Magnetic resonance imaging
NIR	Near infra-red
PIP	Proximal interphalangeal joints
QDA	Quadratic discriminate analysis
RA	Rheumatoid arthritis
RF	Rheumatoid factor
rSQP	Reduced Hessian sequential quadratic programming
Se	Sensitivity
SOM	Self-organizing map
Sp	Specificity
SPN	Simplified spherical harmonics
SPN (CD)	SPN model with complex-valued diffusion operators
SPN (RD)	SPN model with real-valued diffusion operators
SS-DOT	Steady-state diffuse optical tomography
SVM	Support vector machine
TD-DOT	Time-domain diffuse optical tomography
TN	True negative
TNF	Tumor necrosis factor
TP	True positive
US	Ultrasound imaging

Diffuse Optical Tomography

DOT is a tomographic imaging modality, conceptually not unlike x-ray computed tomography (CT), where electromagnetic radiation is directed at an object of interest at multiple locations. The incident photons interact with the media of interest and are either annihilated (absorbed) or scattered. However, unlike x-ray imaging, which uses high energy electromagnetic radiation with wavelengths in the range of 0.01 to 0.1 nm, DOT relies on non-ionizing radiation with wavelengths between about 650 and 900 nm.¹ This window of electromagnetic radiation covers the red side of the visible spectrum and the near infrared (NIR). Conveniently, photons at these wavelengths are non-ionizing. As a result, DOT does not pose the dangers typically associated with radiation dosages due to frequent or prolonged x-ray CT imaging.²

2.1 Absorption and scattering of light in tissue

In DOT, near infrared NIR photons are injected into the imaged object. Photons can be absorbed or scattered by the internal structure of the media. Absorbed photons are annihilated and their energy is converted into thermal energy. Photons that are not absorbed can escape the medium, typically

¹X-rays and γ -rays are energetic enough that their interactions with atoms can result in the ejection of electrons (i.e. “ionizing”) [41]. In contrast, NIR photons are not energetic enough to cause electrons to escape their atoms and therefore their interactions are non-ionizing.

²Thermal damage to tissue can occur if tissue is exposed to an NIR source with sufficient power. However, typical DOT imaging systems all operate at power levels significantly lower than the threshold needed to cause thermal damage and, therefore, exposure to NIR light poses no danger.

after undergoing multiple scattering events.

We return to the comparison between DOT and x-ray imaging to better understand the source of contrast in DOT. X-rays interact directly with atoms, and as a result, the contrast observed in x-ray imaging is generally derived from calcium ($Z = 20$) in bones and other calcifications, iodine ($Z = 53$) in vessels, and oxygen ($Z = 8$) in the lungs [41]. In DOT, contrast is obtained from the interaction of NIR photons with cellular and intra-cellular structures.

Absorption and scattering properties of a tissue constitute the optical properties of the tissue.³ In turn, knowledge of tissue optical properties can be directly exploited for diagnostic and therapeutic purposes. NIR photons are ideal for probing the internal physiology of biological tissue because they experience significant scattering and minimal absorption compared to higher (200 to 600 nm) or lower energy photons ($>1,000$ nm). This phenomena occurs because the absorption spectra of blood, or oxygenated (HbO_2) and deoxygenated hemoglobin (Hb), is minimal within this range (Fig. 2.1).⁴ Imaging within this optical window is, therefore, optimal. In this range, NIR photons can be used to probe significant depths (up to <20.0 cm in certain situations) without experiencing complete annihilation [43]. Photons at lower wavelengths ($<600\text{nm}$) experience significantly more attenuation by blood and, therefore, cannot be used to probe significant tissue depths.

Imaging in the NIR window has the added advantage that the absorption coefficient of water, an important and abundant component of the human body, is significantly lower than blood. Imaging with wavelengths beyond 1,000 nm is extremely difficult because the absorption power of blood rises dramatically at these wavelengths. An additional advantage of imaging within the NIR window is that tissue auto fluorescence is minimal in this range, allowing for improved ability to discern photons of interest [43]. Overall, imaging in the NIR optical window is ideal because excellent signal-to-noise ratios can be obtained in this range.

³For completeness, is important to acknowledge that anisotropy and refractive index are also key components of tissue optical properties. However, these properties are often assumed as constants for a particular tissue type.

⁴The actual absorption spectra of tissue chromophores may vary from person to person and between tissue types, among others. Jacques recently published an excellent review of expected optical properties in biological tissue as a function of wavelength in [42].

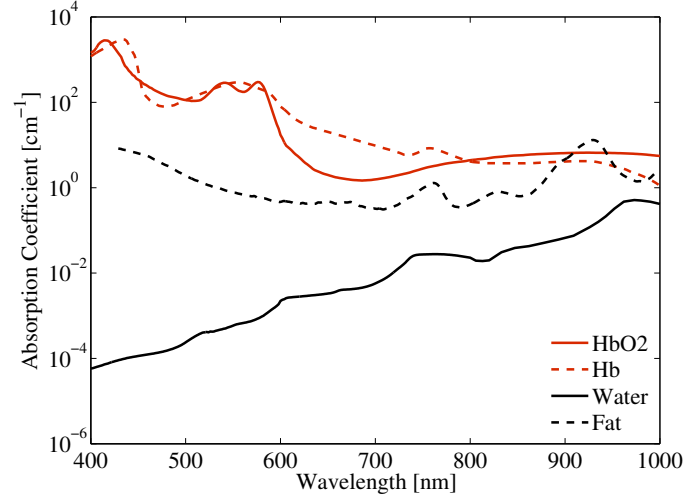


Figure 2.1: Absorption coefficient of oxygenated and deoxygenated hemoglobin (blood), water, and fat in the visible and NIR spectrum. Data for hemoglobin and fat was obtained from [3], citing [4, 5] for hemoglobin data and [6] for fat data. Data for water was obtained from [7], citing [8] as the original source.

The role of internal cellular structures as a source of scattering experienced by light traveling through tissue *ex vivo* was recognized as early as the 1970s, with the reporting of initial studies on scattering characteristics of laser light and the potential to explore inherent cellular scattering properties for diagnostic purposes [44, 45]. Since then, significant effort has been dedicated to understanding which cellular structures are primarily responsible for scattering light. Mismatch in the refractive index between cellular walls and extracellular content is a significant source of scattering [46]. Cell nuclei, mitochondria, and lysosomes have been identified as the primary sources of scattering inside cells [46, 47, 48].

The relative percent of scattering each cellular structure is responsible for varies from cell to cell, as the relative concentration of mitochondria and lysosomes can vary. For illustration purposes, a schematic of a typical animal cell is presented in Fig. 2.2. It is no coincidence that (generally) the structures responsible for scattering NIR light are similar in size to the wavelength of the incident photons, ranging from about 200 nm to 1,000 nm [46, 49]. The topic of light scattering, in general, is a rich and expansive field and beyond the scope of this dissertation. However, an excellent presentation by Bohren and Huffman on the theory of absorption and scattering by small



Figure 2.2: Illustration of a typical mammalian cell. Internal cell structures are labeled: 1) nucleolus, 2) nucleus, 3) ribosome, 4) vesicle, 5) rough endoplasmic reticulum, 6) Golgi apparatus, 7) cytoskeleton, 8) smooth endoplasmic reticulum, 9) mitochondrion, 10) vacuole, 11) cytosol, 12) lysosome, 13) centriole, 14) cell membrane.

particles can be found in [50].

The source of contrast is important, as significant information can be obtained about the state of tissue from its optical properties, as it is well known that cellular composition of tissue changes in the presence of pathologies. For example, anaplasia can be characterized by the presence of enlarged nucleoli and clumped chromatin [51], and can therefore be an important indicator of the presence of malignant neoplasms.

The motivation and indeed, the need, for development of DOT as a viable imaging modality arises from the understanding that DOT is sensitive to physiological processes not accessible by other imaging modalities. Mainly, NIR photons can be used to probe tissue and to determine three-dimensional maps of absorption and scattering coefficients. These maps can then be analyzed to determine the state of the underlying physiology. This information can be used to diagnose disease, monitor physiological processes, or characterize tissue.

2.2 Models of light propagation in tissue

Escaping photons can be detected with contact-based or contact-free techniques, using arrays of photodiodes, photomultiplier tubes (PMT), or an intensified charged-coupled device (ICCD). This general process is demonstrated by Fig. 2.3, where possible locations of surface illumination on a

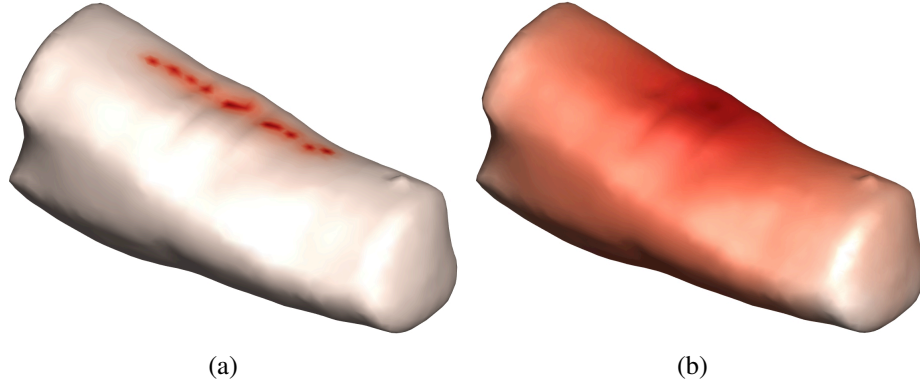


Figure 2.3: (a) Example of laser light illuminating a human finger with NIR photons. (b) A typical distribution of exiting photons at the surface of the finger (i.e. a “red glow”).

human finger and the subsequent wave of photons that escape after undergoing internal scattering are shown.

Contact-based imaging systems typically employ optical fiber cables to deliver photons from a laser source to the imaged object. A separate set of optical fibers collect exiting photons at the surface of the object and deliver them to the detection unit. As the name implies, these optical fibers come into contact with the surface of the imaged object. Contact-free systems typically do not require the use of optical fibers. Instead, the laser beam is typically directly incident on the surface of the object. On the detection side, an ICCD camera and accompanying optical system (i.e. lenses, mirrors, and filters) capture photons exiting the imaged object.

The measured quantity is called the partial current or photon flux, $J^+(\mathbf{r})$, which is a measure of exiting energy with units of photons $\text{s}^{-1} \text{cm}^{-2}$ or W cm^{-2} . The experimentally measured partial current is used in conjunction with a DOT reconstruction algorithm to determine the underlying tissue properties.

The DOT reconstruction process is an inverse problem, where the spatial distribution of optical properties in the medium of interest, denoted as $\mu(\mathbf{r})$ or simply μ , are sought from the experimentally measured partial current or $J_M^+(\mathbf{r})$. There are two components to all DOT algorithms: the forward (\mathcal{F}) and inverse (\mathcal{F}^{-1}) models.

Table 2.1: Definition of variables.

Variable	Description	Units
$J^+(\mathbf{r})$	Partial current	W cm^{-2}
$J_P^+(\mathbf{r})$	Predicted partial current	W cm^{-2}
$J_M^+(\mathbf{r})$	Measured partial current	W cm^{-2}
$S(\mathbf{r}, \boldsymbol{\Omega})$	Boundary source	$\text{W cm}^{-2}\text{sr}^{-1}$
N	Order of SP_N model	
\mathbf{r}	Spatial position	
g	Anisotropy factor	
$\mu(\mathbf{r})$	General inverse variable	
$\mu_t(\mathbf{r})$	Total attenuation coefficient	cm^{-1}
$\mu_a(\mathbf{r})$	Absorption coefficient	cm^{-1}
$\mu_s(\mathbf{r})$	Scattering coefficient	cm^{-1}
$\mu'_s(\mathbf{r})$	Reduced scattering coefficient	cm^{-1}
$\mu_{an}(\mathbf{r})$	n th order absorption coefficient	cm^{-1}
$\psi(\mathbf{r}, \boldsymbol{\Omega})$	Radiance	$\text{W cm}^{-2} \text{sr}^{-1}$
$\varphi_n(\mathbf{r})$	n th order composite moment of the SP_N model	W cm^{-2}
$\phi_n(\mathbf{r})$	n th order Legendre moments of the radiance	W cm^{-2}
$\hat{\mathbf{n}}$	Normal vector at outer surface of \mathbf{O}	
$\boldsymbol{\Omega}$	Angular direction	
v	Speed of light in tissue	cm s^{-1}
ω	Modulation frequency	s^{-1}
$Q(\mathbf{r}, \boldsymbol{\Omega})$	Interior source	$\text{W cm}^{-3}\text{sr}^{-1}$
J_n	Coefficients of the partial current	
$R(\theta)$	Reflectivity	
$p(\boldsymbol{\Omega}, \boldsymbol{\Omega}')$	Scattering phase function	sr^{-1}
n_m	Refractive index of medium (i.e. tissue)	
n_s	Refractive index of outer medium (i.e. air)	
i	Complex unit	
V	Imaging domain	
∂V	Boundary of imaging domain	

$$\begin{aligned}\mathcal{F}[\mu(\mathbf{r}), \mathbf{S}] &= J^+(\mathbf{r}) \\ \mathcal{F}^{-1}[J^+(\mathbf{r}), \mathbf{S}] &= \mu(\mathbf{r})\end{aligned}\tag{2.1}$$

The forward model (\mathcal{F}) is the mathematical construct used to model light propagation in a media that is a function of parameters that influence the propagation of light in the media, including the absorption and scattering, and the source of NIR light (\mathbf{S}). The forward model is used to compute predicted partial current values, or $J_P^+(\mathbf{r})$, as a function of the given set of optical properties and light sources. The inverse model (\mathcal{F}^{-1}) is the mathematical construct that exploits experimental measurements $J_M^+(\mathbf{r})$ to reconstruct the optical properties of tissue, fluorescent probe locations and concentrations, and the locations of bioluminescent probes.

In general, forward models are not directly invertible (at least not in a feasible manner) and an analytical expression for \mathcal{F}^{-1} does not exist. The inversion process is therefore posed as an optimization problem where the forward model is repeatedly evaluated in order to obtain the optical property distribution that minimizes the error between the experimentally measured and simulated partial current, $J_M^+(\mathbf{r})$ and $J_P^+(\mathbf{r})$, respectively.

The optical property variable μ typically includes the absorption (μ_a) and scattering (μ_s) coefficients, but may also include any other parameters in the forward model. In general, the process is iterative, where an initial set of optical properties μ^0 are updated by the inverse model until an acceptable set of optical properties μ^k is reached, for $k \geq 0$. Because the forward model plays such an important role in the reconstruction process, its accuracy and computational complexity are important factors to understand.

There has long been an interest in modeling how light propagates in biological tissues. The emergence of novel diagnostic and therapeutic methods that rely on lasers in the eighties and nineties led to a need for accurately modeling light-tissue interaction. Various models were developed that made use of ever increasing computational power and new numerical methods. Fundamentally two different approaches were pursued: (a) Monte Carlo (MC) modeling [52, 53, 54]

and (b) numerical solutions to the equation of radiative transfer (ERT) and its approximations [17, 19, 20, 55, 56, 57, 58, 59, 60, 61, 62, 63].

The most general method for modeling light propagation is through stochastic simulations of photon-tissue interaction with MC, which employs statistical techniques to propagate a large number of photons through tissue. MC simulations can yield very accurate results and the technique is easy to implement. However, accurate MC solutions require simulating the propagation of a large number of individual photons, typically > 1 Billion. This leads to undesirably long computation times (often days). Recent advances have seen the MC method implemented on massively parallel GPUs, resulting in improvements to the computation time by several orders of magnitude [54, 64]. As a trade off, implementation of MC on GPUs leads to increases in implementation complexity. Overall, the MC method still has only limited utility in DOT.

The choice of forward model has a profound effect on the accuracy of the reconstructed optical maps. In general, more accurate forward models can yield more accurate reconstructions. The ideal choice of forward model is heavily dependent on multiple criteria, including the expected range and distribution of optical properties in the medium. However, the decision in choosing a forward model must strike a balance between the need to model light propagation accurately and the need to perform the required computation in a reasonable amount of time (i.e. minutes or hours instead of days). While highly accurate results can be achieved, MC methods are computationally very intensive. In general, finite-differences, finite-element, and finite-volume methods in which the ERT or one of its approximations is discretized and the resulting system of equations is solved numerically are less computationally expensive. However, the complexity of these codes can still lead to various computational challenges.

The remainder of this chapter is dedicated to reviewing the use of the ERT model in DOT. The formulation of the ERT model in the time, frequency and steady-state domains are reviewed in Section 2.3. Then, a review of the diffusion equation and the simplified spherical harmonics (SP_N) approximations to the ERT are presented in Sections 2.4 and 2.5, respectively. Two specific SP_N formulations of interest are identified in this section. A brief overview of computational grids is

presented in Section 2.6. Chapter 2 concludes with Section 2.7, a discussion and summary of the results presented in this chapter.

2.3 Equation of radiative transfer

The ERT is widely regarded as the most accurate deterministic model for light propagation [19, 20, 58, 59, 60, 61, 62, 63]. The ERT is an energy balance equation where energy propagation is conserved through spatial and angular discretization of the computational domain. Compared to MC, the ERT is extremely difficult to implement because it is an integral-differential equation. The most popular method for solving the ERT is the upwind-step discretization of the differential operator and discrete ordinates (S_N) discretization of the in-scatter (integral) operator. Finite-differences, finite-element, and finite-volume methods in which the ERT is discretized and the resulting system of equations is solved numerically are, generally, computationally demanding, yet less expensive than the MC method.⁵

2.3.1 Time-dependent ERT equation

The most general ERT model (Eq. 2.2) is the system of time-dependent equations that model the radiance $\psi(\mathbf{r}, \boldsymbol{\Omega}, t)$ with units of $\text{W cm}^{-2} \text{ sr}^{-1}$ at each mesh point in the computational domain located at position \mathbf{r} , along discrete angle $\boldsymbol{\Omega}$, at a particular moment in time t , and with a spatially varying internal source $Q(\mathbf{r}, t)$.

$$\left(\frac{1}{v} \frac{\partial}{\partial t} + \boldsymbol{\Omega} \cdot \nabla + \mu_t(\mathbf{r}) \right) \psi(\mathbf{r}, \boldsymbol{\Omega}, t) = \mu_s(\mathbf{r}) \int_{4\pi} p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') \psi(\mathbf{r}, \boldsymbol{\Omega}', t) d\boldsymbol{\Omega}' + \frac{Q(\mathbf{r}, t)}{4\pi} \quad (2.2)$$

The corresponding partially-reflecting boundary equations (Eq. 2.3) are a function of the reflectivity $R(\boldsymbol{\Omega}' \cdot \hat{\mathbf{n}})$ and boundary source $S(\mathbf{r}, \boldsymbol{\Omega}, t)$.

⁵Here, and throughout this text, “expense” refers to the commutative computational resources required to execute a given task. This include the computation time and the memory (RAM) requirements.

$$\psi(\mathbf{r}, \boldsymbol{\Omega}, t) = S(\mathbf{r}, \boldsymbol{\Omega}, t) + R(\boldsymbol{\Omega}' \cdot \hat{\mathbf{n}}) \psi(\mathbf{r}, \boldsymbol{\Omega}', t), \quad \mathbf{r} \in \partial V, \quad \boldsymbol{\Omega} \cdot \hat{\mathbf{n}} < 0 \quad (2.3)$$

The reflectivity corresponds to the fraction of outgoing radiance along $\boldsymbol{\Omega}'$ that specularly reflects into the angular direction $\boldsymbol{\Omega}$. The reflectivity $R(\boldsymbol{\Omega}' \cdot \hat{\mathbf{n}}) = R(\cos \theta)$ is a function of the mismatch in refractive index at the boundary on the imaged object and is defined by Eq. 2.4.

$$R(\cos \theta) = \begin{cases} \frac{1}{2} \left(\frac{n_m \cos \theta' - n_s \cos \theta}{n_m \cos \theta' + n_s \cos \theta} \right)^2 + \frac{1}{2} \left(\frac{n_m \cos \theta - n_s \cos \theta'}{n_m \cos \theta + n_s \cos \theta'} \right)^2, & \text{if } \theta < \theta_c \\ 1, & \text{if } \theta \geq \theta_c \end{cases} \quad (2.4)$$

The reflectivity function is derived from Snell's Law and is a function of the incident angle inside the medium (θ), the refraction angle outside the medium (θ'), the critical angle for total internal reflection θ_c , and the refractive index of the medium (n_m) and outside the medium (n_s). The speed of light in the medium, $v = c/n_m$ is also dependent on the refractive index of the medium. The total attenuation coefficient, μ_t , is defined as

$$\mu_t(\mathbf{r}) = \mu_a(\mathbf{r}) + \mu_s(\mathbf{r}). \quad (2.5)$$

The anisotropic scattering factor, g , acts on μ_s , resulting in the reduced scattering coefficient $\mu'_s = (1-g)\mu_s$. The phase function, $p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}')$, models the highly forward-peaked scattering nature of light propagation in biological media and is typically a function of g .⁶ The Henyey-Greenstein phase function is typically used for biological media,

⁶A detailed discussion of the anisotropic scattering parameter, g , by Bohren and Huffman can be found in [50].

$$p(\cos \theta) = \frac{1 - g^2}{4\pi (1 + g^2 - 2 \cos \theta)^{3/2}}, \quad (2.6)$$

with $g \in [0.8, 1.0]$ for most biological media. The quantity measured by detectors is the partial current, an approximation to the average energy exiting over a given surface area, and is computed from the radiance as follows:

$$J^+(\mathbf{r}, t) = \int_{\Omega \cdot \hat{\mathbf{n}} > 0} [1 - R(\Omega \cdot \hat{\mathbf{n}})] (\Omega \cdot \hat{\mathbf{n}}) \psi(\mathbf{r}, \Omega, t) d\Omega. \quad (2.7)$$

Another quantity of interest is the fluence $\phi(\mathbf{r}, t)$ with units of W cm^{-2} . This quantity will be important when comparing the solutions of the ERT equations to those admitted by the SP_N mode. The fluence is defined as

$$\phi(\mathbf{r}, t) = \int_{4\pi} \psi(\mathbf{r}, \Omega, t) d\Omega. \quad (2.8)$$

2.3.2 Frequency-dependent ERT equation

The frequency-dependent ERT model is derived from the time-dependent ERT model (2.2-2.3) through application of the Fourier Transform, and is given by

$$\left(\Omega \cdot \nabla + \mu_t(\mathbf{r}) + \frac{i\omega}{v} \right) \psi(\mathbf{r}, \Omega, \omega) = \mu_s(\mathbf{r}) \int_{4\pi} p(\Omega \cdot \Omega') \psi(\mathbf{r}, \Omega', \omega) d\Omega' + \frac{Q(\mathbf{r}, \omega)}{4\pi}, \quad (2.9)$$

$$\psi(\mathbf{r}, \Omega, \omega) = S(\mathbf{r}, \Omega, \omega) + R(\Omega' \cdot \hat{\mathbf{n}}) \psi(\mathbf{r}, \Omega', \omega), \quad \mathbf{r} \in \partial V, \quad \Omega \cdot \hat{\mathbf{n}} < 0. \quad (2.10)$$

The imaginary term, $\mu_t + \frac{i\omega}{v}$, is useful in modeling the accumulation of phase within the medium (i.e. offsets in phase between the input signal and the measured signal). The amplitude of the external source, $S(\mathbf{r}, \boldsymbol{\Omega}, \omega)$, is modulated at a known rate, typically 100 to 600 MHz. The modulation frequency (ω) has units of radians. The introduction of an imaginary term in the ERT equations results in a complex attenuation coefficient, $\mu_t + \frac{i\omega}{v}$. The partial current, $J^+(\mathbf{r}, \omega)$, and fluence, $\phi(\mathbf{r}, \omega)$, are defined by (2.7) and (2.8).

2.3.3 Time-independent ERT equation

The time-independent ERT equations (so called “steady-state”) are obtained from the time-dependent or frequency-dependent formulations, by setting the rate of change of the radiance over time to zero, $\frac{\partial}{\partial t}\psi(\mathbf{r}, \boldsymbol{\Omega}, t) = 0$, or by setting the source modulation frequency to zero, $\omega = 0$, resulting in

$$\boldsymbol{\Omega} \cdot \nabla \psi(\mathbf{r}, \boldsymbol{\Omega}) + \mu_t(\mathbf{r}) \psi(\mathbf{r}, \boldsymbol{\Omega}) = \mu_s(\mathbf{r}) \int_{4\pi} p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') \psi(\mathbf{r}, \boldsymbol{\Omega}') d\boldsymbol{\Omega}' + \frac{Q(\mathbf{r})}{4\pi}, \quad (2.11)$$

$$\psi(\mathbf{r}, \boldsymbol{\Omega}) = S(\mathbf{r}, \boldsymbol{\Omega}) + R(\boldsymbol{\Omega}' \cdot \hat{\mathbf{n}}) \psi(\mathbf{r}, \boldsymbol{\Omega}'), \quad \mathbf{r} \in \partial V, \quad \boldsymbol{\Omega}' \cdot \hat{\mathbf{n}} < 0. \quad (2.12)$$

The partial current, $J^+(\mathbf{r})$, and fluence, $\phi(\mathbf{r})$, are defined similar to (2.7) and (2.8).

2.3.4 Numerical challenges associated with the ERT

Widespread use of the ERT in DOT has been limited because numerical solutions to the ERT suffer from two main computational challenges: (1) numerical algorithms for solving the ERT are very difficult to implement and (2) the number of resulting equations from spatial and angular discretization can be prohibitively large, requiring massive system memory and computational power.

There are three main reasons for the high computational cost: 1) the ERT is only first order accurate, and as such, ensuring numerical stability requires the spatial discretization of the compu-

tation domain to be small, therefore requiring the use of highly dense meshes; 2) accurate modeling of highly forward peaked light (as is often the case in tissue) requires high angular discretization to accurately capture the scattering pattern of light in tissue; 3) the presence of scattering makes intensities at different directions strongly coupled, leading to slow convergence. Altogether, the total number of equations that result from the spatial and angular discretization of the ERT is very large, making the computation time required to solve the system of equations impractical, and can often take hours [19, 65, 66].

The differential operator in the ERT is a first order directional derivative. The upwind step-method differencing scheme (SD) is the standard method for solving this type of differential equation [19, 67, 68, 69, 70, 71]. This type of spatial discretization requires the computational domain to be discretized into very small discrete spatial (mesh) elements because the step-method is only first order accurate. Thus, large discretization of the computational domain could easily lead to large numerical errors that can make the numerical solution unreliable. For example, it has been shown that spatial discretization in the order of $\Delta x < 0.001$ cm is required for media with tissue-like optical properties ($\mu_a = 1.0 \text{ cm}^{-1}$ and $\mu'_s = 10.0 \text{ cm}^{-1}$) [1]. Typical tissue volumes are in the order of 2-4 cm per side, requiring over 8 Billion discrete mesh points (in a typical Cartesian grid).

Numerical quadrature (i.e. extended trapezoidal rule) replaces the integral term. The standard unit sphere is discretized into a set of distinct sections that are best represented by a set of directional cosines and corresponding weights. This discretization is called discrete ordinate (S_N), where the corresponding number of discrete ordinates is $\Omega = N(N + 2)$.⁷ Light propagation in tissue experiences extremely forward peaked scattering, and therefore, highly dense angular discretization is required to accurately model the scattering pattern of light. Modeling partially reflective boundaries ($n_m \approx 1.30 - 1.50$ for tissue) also requires high angular discretization because the reflectivity $R(\theta')$ is a strongly varying function when $n_m > 1.0$. In the example given above, consider strong anisotropic scattering ($g > 0.8$) and partially reflective boundaries ($n_m = 1.37$). Accurate modeling of this system requires S_{16} discretization, which corresponds to 288 discrete

⁷Note that the variable N in this formulation differs from the equally-named (perhaps poorly so) variable N that appears in the SP_N formulation.

angular directions.

Thus, in the example, for each of the 8 Billion mesh points there are 288 associated angular directions. Computationally, this implies that a total of 2.3 Trillion data points must be stored. Solving the forward problem with this system requires solving a system of 2.3 Trillion unknowns, which is computationally demanding. The inverse problem requires solving for 8 Billion unknowns, also computationally challenging. This poses severe computational and memory allocation difficulties. Solving for this many unknowns (in the forward problem) is computationally challenging and can take days. This makes the ERT impractical in most applications. However, in practice, the ERT is solved with fewer mesh points and fewer discrete ordinates, which can result in unwanted numerical error that may, in fact, undo any advantage that is gained by using the ERT model instead of one of its approximations.

In addition to computational complexity, the combination of the upwind step-method and discrete ordinates poses algorithm implementation challenges because the “upwind” direction changes according to the “ordinate” being considered. This makes implementation of an ERT-based algorithm difficult.

As a result of these computational considerations, approximations to the ERT have become increasingly important and common. Significant effort has been dedicated to developing approximations to the ERT that yield similar accuracy but at significantly lower computational costs. The diffusion equation (DE) approximation to the ERT is by far the most widely used light propagation model in DOT. The simplified spherical harmonics approximation (SP_N) to the ERT has seen an increase in interest over that past 7 years as it has been shown to be an improvement over the DE with only a moderate increase in computational cost [17, 19].

2.4 Diffusion approximation to the ERT

The approximation to the ERT most commonly used is the DE, which is the P_1 approximation to the ERT [17]. The DE can be solved with minimal computation power, but there are significant

constraints on the type of biological tissue within which it is a valid approximation to the ERT. The DE is computationally attractive because, unlike the ERT model, its solution is not angular dependent, thus reducing the linear equations that result from discretization of the angular variable in (2.2). In addition, the diffusion operator is a Laplacian operator and can be naturally discretized by second-order accurate numerical schemes. The increase in numerical accuracy provided by the second-order spatial discretization allows for the use of a less dense computation mesh because the numerical error is $O(\Delta x^2)$.⁸ The frequency-dependent DE is presented as an example [19].

$$-\nabla \cdot \frac{1}{3[\mu_a + (1-g)\mu_s]} \nabla \phi(\mathbf{r}, \omega) + \left(\mu_a + \frac{i\omega}{v} \right) \phi(\mathbf{r}, \omega) = Q(\mathbf{r}, \omega) \quad (2.13)$$

$$\left(\frac{1}{2} - R_1 \right) \phi(\mathbf{r}, \omega) + \frac{1 + 3R_2}{3[\mu_a + (1-g)\mu_s]} (\hat{\mathbf{n}} \cdot \nabla \phi(\mathbf{r}, \omega)) = \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) 2|\Omega \cdot \hat{\mathbf{n}}| d\Omega \quad (2.14)$$

The meaning of the coefficients R_1 and R_2 will be made clear in Section 2.5. While not explicitly shown, the absorption, scattering, and anisotropy coefficients are spatially varying.

The DE assumes that light propagates through tissue in a diffuse manner. This assumption is valid only under conditions where light is highly scattered and infrequently absorbed ($\mu_a \ll \mu_s$). Furthermore, the media of interest must be geometrically large enough to allow for light (potentially collimated and highly directional) to become sufficiently scattered (i.e. diffuse). Thus, the DE has been shown to be a poor approximation to the ERT in tissue of small volume, with high absorption, void-like regions, or internal light sources near the boundary [65]. This limits the applications for which the DE approximation can be used, making it particularly poor if used to model light propagation in tissue with high absorption (liver, arteries, heart, kidneys, feet), void like regions (brain, lungs), or in small volumetric geometries (mice, finger joints).

The diffusion equation is most applicable in DOT imaging of human female breast tissue. Breast tissue is known to be highly scattering with low absorption properties and very homogeneous. Furthermore, the geometry is typically sufficiently large enough to allow for collimated

⁸This implies that the distance between mesh points can be significantly increase without loss of accuracy. A detailed analysis of numerical error associated with spatial distances between mesh points can be found in [72].

light to become diffuse. However, the diffusion equation is expected to give unreliable results if used for optical tomographic imaging of fingers because the geometries are small, they contain highly absorbing regions (i.e. vasculature), there are internal void like regions (i.e. joint cavity), and the tissue is extremely inhomogeneous (bone, tendons, ligaments, muscle, vasculature, synovium).

2.5 SP_N approximation to the ERT

The simplified spherical harmonics (SP_N) approximation to the ERT has been shown to be superior to the DE at modeling light propagation in tissue of small volume, with higher absorption, and with lower scattering [1]. The SP_N approximation to the ERT in the steady-state domain (time-independent) was introduced to tissue optics in 2006 by Klose and Larsen [1]. Chu et al. introduced initial work on the expansions of the SP_N equations to the frequency-domain in 2009 [73], while Dominguez et al. expanded the theory to the time-domain in 2010 [74]. In the same work, Dominguez et al. also proposed an alternative formulation of the frequency-dependent SP_N model.

Klose and Larsen showed that the SP_N ($N > 1$) approximation to the ERT is superior to the diffusion equation (DE or SP_1) at modeling light propagation in tissue of small volume, with higher absorption, and with lower scattering. The main advantage of the SP_N equations over other deterministic light propagation models is that they are a better approximation to the ERT than the DE and are far less computationally demanding than the ERT. Computational demand is reduced because the number of coupled equations resulting from the SP_N approximation is substantially fewer than those resulting from the ERT. Furthermore, the SP_N equations are second-order accurate, and therefore, the computation mesh does not need to be as dense as the mesh required to accurately solve the ERT. The SP_N approximation to the ERT has been used in the steady-state domain using finite differences discretization [1], in the frequency-domain using finite element discretization [73], and in the time-domain using the finite element method [74].

2.5.1 Time-independent SP_N equations

The formal derivation of the time-independent SP_N model (also know as the steady-state SP_N model or SS- SP_N) by Klose and Larsen is presented with full detail in [1] and are derived from the time-independent ERT model (2.11-2.12). Below, the SP_3 equations are reproduced in detail from [1]. The SP_1 , SP_5 , and SP_7 equations are omitted for brevity. The SP_3 equations model the composite moments φ_n and are defined as

$$\begin{aligned} -\nabla \cdot \frac{1}{3\mu_{a1}} \nabla \varphi_1 + \mu_a \varphi_1 &= Q + \frac{2}{3} \mu_a \varphi_2 \\ -\nabla \cdot \frac{1}{7\mu_{a3}} \nabla \varphi_2 + \left(\frac{4}{9} \mu_a + \frac{5}{9} \mu_{a2} \right) \varphi_2 &= -\frac{2}{3} Q + \frac{2}{3} \mu_a \varphi_1. \end{aligned} \quad (2.15)$$

The corresponding equations for modeling partially reflecting boundaries are the following,

$$\begin{aligned} \left(\frac{1}{2} + A_1 \right) \varphi_1 + \left(\frac{1 + B_1}{3\mu_{a1}} \right) \hat{\mathbf{n}} \cdot \nabla \varphi_1 &= \left(\frac{1}{8} + C_1 \right) \varphi_2 + \left(\frac{D_1}{\mu_{a3}} \right) \hat{\mathbf{n}} \cdot \nabla \varphi_2 \\ &+ \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) 2|\Omega \cdot \hat{\mathbf{n}}| d\Omega, \\ \left(\frac{7}{24} + A_2 \right) \varphi_2 + \left(\frac{1 + B_2}{7\mu_{a3}} \right) \hat{\mathbf{n}} \cdot \nabla \varphi_2 &= \left(\frac{1}{8} + C_2 \right) \varphi_1 + \left(\frac{D_2}{\mu_{a1}} \right) \hat{\mathbf{n}} \cdot \nabla \varphi_1 \\ &+ \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) (5|\Omega \cdot \hat{\mathbf{n}}|^3 - 3|\Omega \cdot \hat{\mathbf{n}}|) d\Omega. \end{aligned} \quad (2.16)$$

In (2.16), $S(\Omega)$ denotes the boundary source, Ω represents the angular discretization of boundary points, and $Q(\Omega)$ is a volumetric source. The coefficients A_i , B_i , C_i , and D_i are functions of various angular moments of the boundary reflectivity [1]. The variables of interest φ_i are the composite moments of the fluence, and the Legendre moments (or fluence) ϕ_i can be written as functions of the composite moments listed below,

$$\begin{aligned}\phi_2 &= \frac{1}{3}\varphi_2, \\ \phi_0 &= \varphi_1 - \frac{2}{3}\varphi_2.\end{aligned}$$

The corresponding partial current equation is

$$\begin{aligned}J^+(\mathbf{r}) &= \left(\frac{1}{4} + J_0\right) \left(\varphi_1 - \frac{2}{3}\varphi_2\right) - \left(\frac{0.5 + J_1}{3\mu_{a1}}\right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) \\ &\quad + \left(\frac{5}{16} + J_2\right) \left(\frac{1}{3}\varphi_2\right) - \left(\frac{J_3}{7\mu_{a3}}\right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2).\end{aligned}\tag{2.17}$$

The boundary coefficients $A_1, B_1, C_1, D_1, A_2, B_2, C_2,$ and D_2 and partial current coefficients $J_0, J_1, J_2,$ and J_3 are defined in Table 2.2 for the SP_3 model. The coefficients for higher order models can be found in [1]. The n th order absorption coefficients are a function of $\mu_a, \mu_s, g,$ and the order of the SP_N model (2.18).

$$\mu_{an} = \mu_t - g^n \mu_s \tag{2.18}$$

The boundary and partial current coefficients are all functions of various angular moments of the boundary reflectivity R_n , which are computed as a function of the reflectivity $R(\Omega)$ and the discrete angle Ω (typically obtained using the discrete ordinates method) using the following equation,

$$R_n = \int_0^1 R(\Omega) \Omega^n d\Omega \tag{2.19}$$

Table 2.2: Boundary and current coefficients for the SP₃ model as defined by Klose and Larsen in [1].

$$\begin{aligned}
A_1 &= -R_1 & J_0 &= -\frac{1}{2}R_1 \\
B_1 &= 3R_2 & J_1 &= -\frac{3}{2}R_2 \\
C_1 &= -\frac{3}{2}R_1 + \frac{5}{2}R_3 & J_2 &= \frac{5}{4}R_1 - \frac{15}{4}R_3 \\
D_1 &= \frac{3}{2}R_2 - \frac{5}{2}R_4 & J_3 &= \frac{21}{4}R_2 - \frac{35}{4}R_4 \\
A_2 &= -\frac{9}{4}R_1 + \frac{15}{2}R_3 - \frac{25}{4}R_5 & \phi_0 &= \varphi_1 - \frac{2}{3}\varphi_2 \\
B_2 &= \frac{63}{4}R_2 - \frac{105}{2}R_4 + \frac{175}{4}R_6 & \phi_1 &= \frac{1}{3}\varphi_2 \\
C_2 &= -\frac{3}{2}R_1 + \frac{5}{2}R_3 \\
D_2 &= \frac{3}{2}R_2 - \frac{5}{2}R_4
\end{aligned}$$

2.5.2 Time-dependent SP_N equations

The SP_N system in the time-domain were developed by Dominguez and Bérubé-Lauzière in [74]. The derivation of the time-dependent parabolic (TD-pSP_N) equations takes into account ray-divergence effects. The derivation of TD-pSP_N also follows the derivation method presented by Klose and Larsen in [1] for the steady-state SP_N equations. An important difference results from the inclusion of the time-dependent term in the transport equations (2.2). The n th order absorption coefficients differ from the steady-state version only in the inclusion of a “divergence coefficient” $\mu_d(\mathbf{r})$ that occurs as a results of the chosen ERT model, which differs from (2.2) in the inclusion of spatially varying refractive index $\eta(\mathbf{r})$ and a term that models ray divergence. The n th order absorption coefficients are defined as

$$\mu_{an} = \mu_a + \mu_s(\mathbf{r}) [1 - g(\mathbf{r})^n] + \mu_d(1 - \delta_{n,0}), \quad (2.20)$$

where $\delta_{n,0}$ is the Kronecker delta. By assuming a spatially invariant refractive index and neglecting

the effects of ray divergence, we reduce the n th order absorption coefficients to (2.18) as presented by Klose and Larsen in [1]. The TD-SP $_N$ for $N = 3$ are reproduced below for comparison with the steady-state model in Section 2.5.1.

$$\begin{aligned} -\nabla \cdot \frac{1}{3\mu_{a1}} \nabla \varphi_1 + \left(\mu_a + \frac{1}{v} \frac{\partial}{\partial t} \right) \varphi_1 &= Q + \frac{2}{3} \left(\mu_a + \frac{1}{v} \frac{\partial}{\partial t} \right) \varphi_2 \\ -\nabla \cdot \frac{1}{7\mu_{a3}} \nabla \varphi_2 + \left(\frac{4}{9}\mu_a + \frac{5}{9}\mu_{a2} + \frac{1}{3} \frac{1}{v} \frac{\partial}{\partial t} \right) \varphi_2 &= -\frac{2}{3}Q + \frac{2}{3}\mu_a \varphi_1 \end{aligned} \quad (2.21)$$

If we neglect the divergence coefficient μ_d in the n th order absorption coefficients for the TD-SP $_N$ (2.20), then the boundary equations of the TD-SP $_N$ model are almost identical to the boundary equations of the SS-SP $_N$ model in (2.16). The only difference is that the composite moments $\varphi_n(\mathbf{r}, t)$ and boundary source $S(\mathbf{r}, t)$ are functions of space and time.

2.5.3 Frequency-dependent SP $_N$ equations

Dominguez and Bérubé-Lauzière [74] convert the TD-SP $_N$ model to a version of the frequency-domain SP $_N$ (FD-SP $_N$) model by application of the Fourier Transform. The resulting system of equations is similar to the TD-SP $_N$, the only difference being the replacement of the time-dependent term $\frac{\partial}{\partial t}$ with an imaginary term that depends on the source modulation frequency ω .

FD-SP $_N$ Model 1: Real-valued diffusion coefficient

$$\begin{aligned} -\nabla \cdot \frac{1}{3\mu_{a1}} \nabla \varphi_1 + \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 &= Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_2 \\ -\nabla \cdot \frac{1}{7\mu_{a3}} \nabla \varphi_2 + \left(\frac{4}{9}\mu_a + \frac{5}{9}\mu_{a2} + \frac{1}{3} \frac{i\omega}{v} \right) \varphi_2 &= -\frac{2}{3}Q + \frac{2}{3}\mu_a \varphi_1 \end{aligned} \quad (2.22)$$

As in the case of the TD-SP $_N$ model, if we neglect the divergence coefficient in (2.20), then the term μ_{an} in (2.22) is identical to the definition used in the SS-SP $_N$ model. The resulting boundary conditions and partial current equations are therefore also identical. The FD-SP $_N$ equations solve

for complex valued composite moments $\varphi_n(\mathbf{r}, \omega)$ using complex valued sources, $Q(\mathbf{r}, \omega)$ and $S(\mathbf{r}, \omega)$.

The FD-SP_N equations are characterized by the inclusion of the imaginary frequency component $\frac{i\omega}{v}$, where ω represents the frequency with which the amplitude of the source is modulated (in radians), v is the speed on light inside the medium (accounting for the refractive index n_m of the medium and the speed of light in a vacuum c), and i is the imaginary unit denoting the imaginary part of a complex number. In the FD-SP_N formulation, the fluence (φ_i) and the source terms (Q and S) become complex-valued (i.e. $\varphi_i = (\varphi_i)_R + i(\varphi_i)_I$).

An alternative FD-SP_N model is presented by Chu et al. in [73], where the FD-SP_N equations are derived through a simple modification of the time-independent (or SS-SP_N) model derived by Klose and Larsen in [1] and summarized in Section 2.5.1. In [73], the authors justify their version of FD-SP_N model by arguing that the “. . . complex nature of the attenuation coefficient arises from the fact that radiance will also vary due to the modulation frequency, with the effect of making the absorption coefficient complex.” In [73], the FD-SP_N equations are derived through a simple modification of the time-independent (or SS-SP_N) model derived by Klose and Larsen in [1] and summarized in Section 2.5.1. The only change occurs in (2.18), by redefining the n th order absorption coefficients to include the imaginary term associated with the source modulation frequency that appears in the FD-ERT model (2.9). The complex-valued n th order absorption coefficients, denoted by $\tilde{\mu}_{an}$, are given by

$$\tilde{\mu}_{an} = \mu_{an} + \frac{i\omega}{v}. \quad (2.23)$$

The authors justify the addition of the imaginary term is by considering that the radiance varies due to modulation frequency.

This same formulation can be derived by considering the one-dimensional FD-ERT model, expanding that model into its P_N approximation, and then following the steps presented by Klose and

Larsen in [1]. This derivation, while tedious, is straight forward by following the steps presented in Section 3.1 of [1]. The only change that must be made to the FD-ERT model is in the definition of the total attenuation coefficient, which changes from $\mu_t(x) = \mu_a + \mu_s$ to $\mu_t(x) = \mu_a + \mu_s + \frac{i\omega}{c}$. Then, carrying that new complex term throughout the rest of the derivation results in a complex-valued n th order absorption coefficient, $\tilde{\mu}_{an} = \mu_a + (1 - g^n)\mu_s + \frac{i\omega}{c}$.

Crucially, deriving the FD-SP_N model from the FD-ERT model results in the appearance of an imaginary term remains in the denominator of the “diffusion” operators and cannot be removed (with any justification).

The FD-SP_N equations can be obtained by inserting (2.23) into the SS-SP_N model in (2.15-2.17). The resulting system of equations is given below.

FD-SP_N Model 2: Complex-valued diffusion coefficient

$$\begin{aligned} -\nabla \cdot \frac{1}{3\tilde{\mu}_{a1}} \nabla \varphi_1 + \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 &= Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_2 \\ -\nabla \cdot \frac{1}{7\tilde{\mu}_{a3}} \nabla \varphi_2 + \left(\frac{4}{9}\mu_a + \frac{5}{9}\mu_{a2} + \frac{i\omega}{v} \right) \varphi_2 &= -\frac{2}{3}Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 \end{aligned} \quad (2.24)$$

This addition of the imaginary term to the definition of the n th order absorption coefficients results in complex “diffusion-like” coefficients. It is noteworthy to point out that this small change is significant for two reasons. First, the magnitude of the complex term is such that $\mu_a \approx \frac{\omega}{v}$ for clinically relevant frequencies (i.e. 300 – 600 MHz). Thus, the modulation frequency term affects the diffusion-like operators to the same extent as the absorption coefficient. Second, the FD-SP₁ equation no longer equals the frequency-domain diffusion equation (FD-DE) approximation to the ERT, as the FD-DE does not have a complex valued diffusion coefficient. Instead, the imaginary term in the FD-DE appears only outside the diffusion coefficient and serves only as an attenuation term (2.13).

The differences between the approach by Dominguez in deriving (2.22) and Chu in deriving (2.24) affect the computation of numerical solutions to these models. The system of equations in (2.22) has complex valued terms only along the main diagonal, whereas the resulting system

of equations from (2.24) has complex-valued coefficients at every entry of the forward model. Numerical solutions to the FD-SP_N models are discussed in more detail in Section 8.3.

2.6 Structured and unstructured computational grids

A major consideration in implementing DOT algorithms that model light-propagation with deterministic models (i.e. non-MC methods) is the type of computational grid used in these numerical calculations. Light transport models are typically solved on either a structured or an unstructured grid. Structured and unstructured grids differ from one another by the method in which Euclidean space is discretized. Structured grids discretize Euclidean space into a set of hexahedron elements (quadrilaterals in two-dimensions), while typical unstructured grids discretize Euclidean space into a collection of tetrahedral elements (triangles in two-dimensions). The connectivity between nodes in a structured grid is implicitly known; nodes always connect to form a cuboid of constant size and orientation. This *a priori* information is central to the finite-differences method, where any given grid point is assumed to have neighboring grid points that also connect to form cuboids. Conversely, the connectivity of unstructured grid elements must be explicitly provided because unstructured grid nodes connect to form elements that vary in size, shape, and orientation [75]. Thus, algorithms that solve the ERT on unstructured grids are more complex than algorithms on structured grids because node connectivity information must be explicitly provided and processed. Additionally, generating unstructured grids can in its-self be an arduous task, often requiring third party applications, while generating structured grids is a relatively simple task.

Unstructured grid methods, however, are well suited for working with complex geometries because the mesh can accurately resolve the physical boundary, while single structured grids cannot. Nonetheless, structured grid methods are attractive because algorithm complexity is minimized as a result of the well-ordered nature of Cartesian grids. Numerical algorithms on both unstructured and structured grids have been developed for solving the ERT. For example, unstructured finite-element methods have been used by Arridge et al. [57], Salah et al. [63], and Rasmussen et

al. [76]. Kim and Hielscher [20], Ren et al. [62], and Gu et al. [77] have employed unstructured finite-volume methods. Finite-differences ERT codes on structured grids have been mainly pursued by Klose et al. for rectangular [19, 61, 67, 78] and irregular [79] geometries. The latter uses a blocking-off method for approximating the physical boundary on a single structured grid.

A detailed discussion of the finite volume method (FVM) is presented in Section 8.1, while a structured- and semi-structured grids are discussed in Appendix C.

2.7 Discussion

The computational costs associated with solving the inverse problem in DOT are a considerable hindrance to more widespread use of this imaging technology. An aim of the work presented in this dissertation is to improve the computational efficiency and speed of DOT algorithms through implementation of novel numerical techniques. This is achieved through two approaches.

First, a DOT algorithm based on the simplified spherical harmonics (SP_N) approximation to the ERT is developed using the finite-volume (FV) discretization method. In particular, the work focuses on using the SP_N equations in the frequency-domain (FV-FD- SP_N) as the forward model. Then, a powerful approach known as PDE-constrained optimization is used as the inverse model, leading to a significant reduction in the reconstruction time. Two algorithms based on the FV discretization of the two competing FD- SP_N models, (2.22) and (2.24), are developed. The performance of each algorithm is evaluated and compared to benchmark solutions obtained with the frequency-dependent ERT equations. This forward model is presented in Chapter 8. The PDE-constrained reconstruction algorithm for the SP_N model is presented in Chapter 9.

Then, the PDE-constrained FV-FD- SP_N algorithm is applied to clinical data.⁹ The resulting DOT images are used to evaluate the ability to diagnose RA using SP_N based images. The application of the PDE-constrained reconstruction algorithm to clinical data is presented in Chapter 10.

used to reconstruct clinical data obtained from a clinical study that imaged the proximal inter-

⁹The clinical data is presented in detail in Section 5.2. It consists of DOT scans of proximal interphalangeal (PIP) joints of 33 subjects (total of 99 joints) with RA and 20 healthy control subjects (total of 120 joints).

phalangeal (PIP) joints of 33 subjects with RA and 20 healthy control subjects. The resulting DOT images are used to diagnose the presence of RA in each of the 219 imaged joints.

Second, the ERT is solved on block-structured grids (BSG), which are a type of semi-structured grids that allow for refinement of the computation mesh near object boundaries without sacrificing the regular structure that is inherent to Cartesian grids. Solving the ERT on structured grids is attractive because algorithm complexity is minimized as a result of the well-ordered nature of Cartesian grids. The BSG algorithm is developed as a method for solving the ERT model at reduced computational burden. This work, while interesting, is not central to the main thesis of this dissertation. It can be found in Appendix C.

Together, these approaches improve the viability of DOT as a clinically relevant imaging technology by reducing the associated computation time, leading to decreased waiting times between data acquisition and image reconstruction.

Rheumatoid Arthritis

To understand the potential benefits of using DOT as a tool for the diagnosis of RA, it is important to first understand the current state of the epidemiology of RA, including the pathogenesis, symptoms, and current treatment guidelines. The focus of this chapter is on providing a review of RA, including the pathogenesis of the disease, current diagnostic criteria, and treatment protocols. Because of the importance of medical imaging to this thesis, a review of the role medical imaging and image analysis play in the diagnostic process is presented.

3.1 Epidemiology

RA is a chronic, progressive, systemic, inflammatory autoimmune disorder that causes chronic inflammation of the synovial membrane of small and large joints [39, 80, 81, 82, 83]. While RA can be mild, it can also be severe, with studies showing that up to 10% of individuals suffering from RA can experience total disability [84]. Several inflammatory cascades, all leading towards persistent synovial inflammation and damage to the associated articular cartilage and underlying bone [85], characterize RA. The disease primarily attacks peripheral joints and their surrounding tendons and ligaments [86].

The etiology of RA is unknown, however, it is the most common inflammatory arthritis [82]. RA is associated with significant pain and disability, affecting about 1% of the population world-

wide, and approximately 1.3-2.1 million people in the US [87,88,89,90,91]. In the US alone, RA leads to 9 million physician visits per year [82]. Women are about three to four times as likely as men to develop RA [92,93]. RA can occur at any age, although it is more likely to occur among those aged 40-70 years and its incidence rate increases with age.

Subjects with RA can suffer from severe pain, joint stiffness, swelling of multiple joints, and lack of joint mobility. Diagnosis of RA at onset is a challenging proposition because the associated clinical features are broad and onset is generally subtle [91]. Symptoms primarily affect the diarthrodial joints, with common examples being the wrists, proximal interphalangeal (PIP), and the metatarsophalangeal joints [82,83,91]. Uncontrolled, the symptoms can lead to self-limiting arthritis or rapidly progressing multi-system inflammation with significant morbidity and mortality (including cardiac, neurological, and hematological complications) [91]. These symptoms can eventually lead to severe disabilities and loss in quality of life [39].

These handicaps can result in large financial costs due to health care expenses and loss of productivity at work. Despite recent advances in therapeutic intervention including biological therapies, there is currently no cure for RA. Early treatment of RA, however, has been shown to significantly improve clinical outcome and management of the disease. It is, therefore, important to diagnose a subject with RA as early as possible.

RA is generally thought to occur due to an overactive immune system. The cytokines tumor necrosis factor (TNF) and interleukin (IL)-1 are believed to play a pivotal role in the pathologic processes of RA. TNF (formerly known as TNF- α) is a monocyte-derived cytotoxin that is synthesized by macrophages in response to proinflammatory stimuli, acting as a central mediator of inflammation and immune regulation. Similar to TNF, IL-1 is a proinflammatory cytokine secreted by macrophages, which acts as a crucial mediator of the immune and inflammatory responses [94]. Overproduction and over-expression of TNF is one of several key inflammatory cascades in RA, and is known to drive synovial inflammation and joint destruction [85].

Synovitis is a key characteristic of RA and is directly linked to multiple alterations to the underlying joint physiology. The differences between a healthy synovium and an inflamed synovium

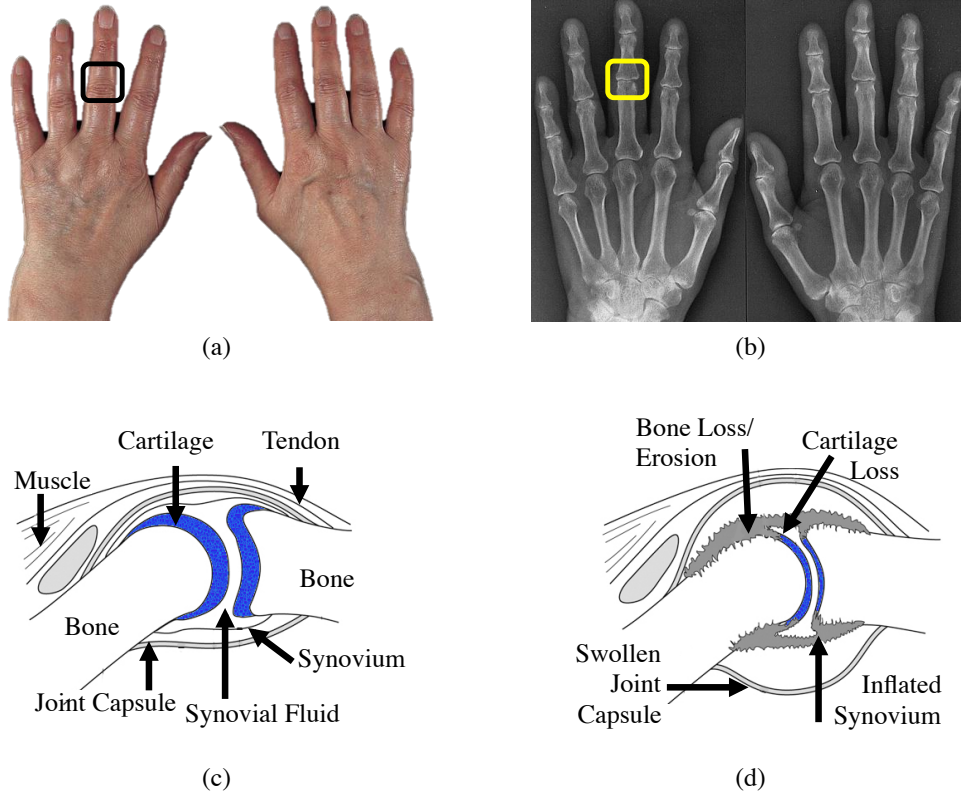


Figure 3.1: (a) Example of symmetric synovitis affecting the hands, highlighting the location of a PIP joint. (b) X-ray images with radiographic evidence of RA, highlighting the location of a PIP joint. (c) Illustration of a typical healthy joint. (d) Illustration of a typical joint exhibiting symptoms of RA. (c,d) Schematic of a typical synovial joint. These joints are surrounded by a “joint capsule” that provides protection, the inside of which is called the synovial membrane or synovium. The synovial capsule is filled with synovial fluid, a normally clear and colorless viscous non-Newtonian fluid whose primary purpose is to provide lubrication to the articulating surface of the joint. The onset of symptoms from RA causes the synovium to become inflamed, causing warmth, redness, swelling, and pain. Progression of the disease can lead to invasion of the inflamed synovium and result in damage to the articulating surfaces and surrounding bone. The synovial fluid undergoes substantial changes, resulting in yellow to red color and cloudy to opaque clarity. Image courtesy of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (<http://images.niams.nih.gov/detail.cfm?id=74> - last accessed June 13, 2012).

are strikingly clear from histological analysis. The inflamed synovium is characterized by pronounced angiogenesis, an influx of inflammatory leucocytes, cellular hyperplasia, and changes in the expression of cell-surface adhesion molecules, proteinases, proteinase inhibitors, and many cytokines [91]. Within a short period from disease onset, the synovium can become hyperplastic, resulting in a thickening of the synovium membrane lining, reaching ten or more cells in depth.

The synovium sublining also undergoes alterations, resulting in prominent infiltration of mononuclear cells (T-cells, B-cells, macrophages, and plasma cells). Multiple other histological changes occur that also alter the properties of the synovium, including elevated T-cell counts in the synovium lining, increased concentration of cytokines (i.e. TNF and IL-1), and elevated expression of adhesion molecules [91].

Furthermore, angiogenesis is over active in RA, particularly during the early onset of the disease, as the newly formed vessels provide oxygen and nutrients to the hypertrophic synovium. The synovium and synovial fluid of subjects with RA can contain a growing list of angiogenic factors that includes cytokines, growth factors, colony-stimulating factors, and soluble adhesion molecules [91].

In its early stage, RA is characterized by inflammatory synovitis that leads to edema in the synovial membrane (synovium) and joint capsule. The permeability of the synovium is changed, leading to an increase of fluid and large cells in the synovial cavity. Subsequently, thick layers of granulation tissue cover and invade the articular cartilage. The joint cartilage and adjacent bone are destroyed. Eventually the joint space closes and muscle and bone undergo atrophy and misalign, leading to subluxation and visible deformity. Fibrous tissue calcifies and leads to immobility of the joint. Different authors have provided classifications of the disease progression into different stages. For example, Steinbrocker differentiates between 4 stages [95,96], while more recently Harris, recognizing the importance of the early pathophysiological changes, divides the progression of RA into 5 stages [88].

3.2 Diagnosis

There is no established pathognomonic feature that can be used to diagnose RA beyond doubt (either biological, clinical, or radiological) [97]. Rather, physicians diagnose RA using a combination of available evidence, including biological, clinical, and radiological information. The lack of a pathognomonic feature for the presence of RA could be a reason for the observed lag between the

onset of symptoms and the clinical diagnosis and subsequent treatment of RA [98].

In practice, the diagnosis of RA is based on the subject's history, physical exam, radiographs, and laboratory studies. The American College of Rheumatology (ACR) has recommended criteria for the classification of RA (the so-called "ACR 1987 revised criteria"). A subject is said to have RA if he or she has satisfied at least four of the seven criteria (Table 3.1). Criteria 1 through 4 must be present for at least 6 weeks.

At the time of publication of the ACR 1987 revised criteria, the sensitivity and specificity of the criteria for classification of RA was 91.2% and 89.3%, respectively [2]. More recently, it was been reported from a systemic literature review that the pooled sensitivity and specificity of the 1987 ACR criteria for diagnosis of established RA were 90% and 80%, respectively. In early RA, the sensitivity and specificity of the criteria were 77% and 77%, respectively. Thus, the criteria can be a poor diagnosis tool for early RA [39]. An updated set of criteria for the diagnosis of RA was reported in 2010 [38]. The 2010 criteria for classification of RA were proposed to enhance the 1987 criteria and to increase the sensitivity of classification of early RA.

Critically, a subject must have obvious "clinical" synovitis present in at least 1 joint before they can be fully assessed and evaluated based on the criteria. It is expected that subjects that met

Table 3.1: Definitions of the ACR 1987 revised criteria for classification of RA [2].

1	Morning stiffness	Morning stiffness in and around the joints, lasting at least 1 hour before maximal improvement.
2	Arthritis of 3 or more joint areas	At least 3 joint areas simultaneously have had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas are right or left PIP, MCP, wrist, elbow, knee, ankle, and MTP joints.
3	Arthritis of hand joints	At least 1 area swollen (as defined above) in a wrist, MCP, or PIP joint.
4	Symmetric arthritis	Simultaneous involvement of the same joint areas (as defined in 2) on both sides of the body (bilateral involvement of PIPs, MCPs, or MTPs is acceptable without absolute symmetry).
5	Rheumatoid nodules	Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxtaarticular regions, observed by a physician.
6	Serum rheumatoid factor (RF)	Demonstration of abnormal amounts of serum rheumatoid factor by any method for which the result has been positive in < 5% of normal control subjects.
7	Radiographic changes	Radiographic changes typical of RA on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized in or most marked adjacent to the involved joints (osteoarthritis changes alone do not qualify).

the 1987 criteria will, very likely, also meet the new criteria for classification of RA. The 2010 criteria are more flexible to the dynamic activity of the disease, allowing for changes in disease classification over time [38].

Early diagnosis of RA is particularly difficult when subjects experience non-erosive symptoms (cannot be detected by radiography, sonography, or MRI scans) and in the absence of the rheumatoid factor (RF) and anti-citrullinated protein (CCP) antibodies [99]. There is no single test that can be used to diagnose, with absolute certainty, the presence of RA. Hence, a combination of physician evaluation, medical imaging, and blood tests are used.

3.3 Treatment

There is no known cure for RA [100]. Recent years have seen substantial advances in the understanding of the pathogenesis and pathophysiology and with it the pharmacotherapy of RA has changed considerably [88, 101, 102, 103, 104, 105, 106]. Recent advances in therapeutic interventions, including biological therapies [107], have increased the ability of physicians to effectively manage subjects with RA when the disease is diagnosed in its early stages.

Many pro-inflammatory cytokines, chemokines and growth factors are expressed in diseased joints, and the recognition of the key role of TNF- α led to the development of highly effective new therapies. TNF- α inhibitors have demonstrated efficacy in clinical trials. It is now clear that the TNF- α blockade, in addition to reducing joint inflammation and leukocyte infiltration, also results in decreased formation of new blood vessels in the synovium. It was shown that anti-TNF- α agents may inhibit joint damage. In the past, so-called disease modifying antirheumatic drugs (DMARD), such as methotrexate, hydroxychloroquine, or sulfasalazine, were initiated only after radiographic evidence of joint damage.

Nowadays, DMARDs are subscribed at much earlier stages in the disease, as it may be possible to prevent or at least delay joint destruction. Because of the poor prognosis of RA and the effectiveness of early treatment, it has become important to diagnose RA as early as possible [39, 82, 83].

Ultimately, the therapeutic goal is to attain sustained remissions.

Clinical studies showed that methotrexate slowed the rate of joint destruction as measured by radiography, and improved quality of life, while separate studies showed superior ability to slow joint erosions with combination therapies that consist of concurrent use of several DMARD agents [108]. More recently, clinical evidence suggests that combination therapies involving DMARDS and biologics results in even better clinical outcomes [109].

Recent pharmacological advances saw the introduction of the first set of biological agents that target TNF- α and are approved for use by subjects with RA. These agents include etanercept (Enbrel), infliximab (Remicade), and adalimumab (Humira). More recent advances have included the addition of new biologics, including the interleukin (IL)-1 receptor antagonist anakinra (Kineret), the T-cell costimulation inhibitor abatacept (Orencia), a monoclonal antibody directed against the B- cell-specific antigen CD20 named rituximab (Rituxan and MabThera), and the anti-TNF- α golimumab (Simponi). In particular, etanercept, a recombinant form of the p75 TNF receptor (TNF-RII), is dimerised via fusion with a portion of the human IgG1 Fc tail, and binds to both TNF- α and TNF- β . Unlike etanercept, infliximab is a humanized mouse monoclonal antibody only against TNF- α [91].

In general, these agents target key components of the host immunodefense system (i.e., TNF- α , IL-1, B-cells, and T-cells). A side of effect of treatment with these biologics is a weakened immune system, which may lead to increased risk of infections that include serious and non-serious bacterial and viral infections [94].

Given the current state of the treatment and the diagnosis, two major aspects concerning improvement have been identified by several researchers in the field. Weinblatt [106] asks for increased efforts concerning early detection so that DMARD therapies can be started earlier. Harris [88] and Jaffe [110] suggest that early detection of the disease and a reliable way to monitor improvement of deterioration would significantly help management of the disease, and Van de Putte [111] notes: “to take optimal advantage of these improved treatment modalities, it now becomes mandatory to further develop relevant measures for monitoring the disease process.”

3.4 The role of laboratory studies

Certain laboratory studies can be helpful in diagnosing inflammatory arthritis, however, they are not specific to RA. The presence of biological markers, such as RF and anti-CCP antibodies, has been used in diagnosis of RA. Tests for RF using a cytofluorimetric assay method results in 80% sensitivity and 87% specificity, respectively [112].

The presence of elevated titers of RF, antibodies to IgG, has been associated with a more severe disease course. RF is increased in 60-85% of subjects with RA. Elevated RF levels can also be observed in cases of sarcoidosis, systemic lupus erythematosus, endocarditis, chronic liver disease, and tuberculosis, among many other conditions [111]. The erythrocyte sedimentation rate (ESR),¹ a measure of the rate at which red blood cells settle, has been found to be elevated in 85-90% of subjects with RA. The ESR is generally a useful measure for following the course of inflammatory activity in an individual subject. Concentration levels of the C-reactive protein (CRP), an acute-phase reactant, may also be used to monitor the level of inflammation. Synovial fluid analysis reveals increased volume and turbidity, but decreased viscosity. The white blood cell count in the synovial fluid commonly exceeds 10,000/mL (synovial-fluid leukocytosis), however, since there is only little fluid in PIP joints, synovial fluid is rarely drawn for analysis.

The role of medical imaging

Of all imaging modalities, X-ray imaging (radiography) has the best-established role in the assessment of the severity of RA [113]. Radiography can document bone damage (erosion) that results from RA and visualize the narrowing of cartilage spaces. However, it has long been recognized that radiography is insensitive to the early manifestations of RA, namely effusion and hypertrophy of the synovial membrane. As a result, x-ray imaging has a limited role in the diagnosis and monitoring of RA, particularly in the early stages of the disease.

The role of ultrasound imaging (US) and magnetic resonance imaging (MRI) in the detection and diagnosis of RA has been a topic of debate [114]. In recent years, US has emerged as a poten-

¹ESR is defined as the rate at which red blood cells sediment over a period of one hour

tially useful imaging technique for diagnosing RA, as it appears to be sensitive to joint effusions and synovial hypertrophy [114, 115, 116, 117, 118].

US has been shown to be a reliable imaging technique for detection of joint erosions, with a particular advantage over radiography and computed tomography in application to early RA [119]. The use of power Doppler US images to aid in diagnosis of RA has led to sensitivity and specificity of 71.1% and 81.8%, respectively. The clinical utility of this technique has yet to be established, as examinations can be tedious and images difficult to interpret. Moreover, US is limited in that it is subject to operator error, which can have important implications as an “acoustic window” is necessary to properly assess the joint structure [119].

MRI is most useful in assessing soft tissue problems, avascular necrosis, degree of cartilage erosion, osteonecrosis and carpal tunnel syndrome [120, 121, 122, 123, 124]. Cumbersome use (long data-acquisition times during which the subject needs to be immobilized), large costs, and the need for contrast agents (e.g. gadolinium to detect increase blood volume caused by neovascularization in the hypertrophic synovial membrane) have prevented MRI from becoming a widely used imaging modality for detection of RA.

Thus, the role of MRI in the diagnosis of RA is more difficult to assess. It has been shown that subjects with early RA (less than two years since onset) cannot be identified from non-contrast MRI scans of finger and wrist joints [125]. The ability to accurately diagnose RA from contrast enhanced MRI (T1 fat saturation with gadolinium injection) scans of the hand showed that sensitivity and specificity of 70% and 64% could be achieved [99]. In a separate study, contrast-enhanced MRI was used to achieve sensitivity and specificity of 100% and 78%, respectively [126].

The widespread use of contrast-enhanced MRI for detection of RA is limited not only for cost and logistical reasons, but also because the use of contrast agents (such as gadolinium) can be difficult to justify due to the high prevalence of nephropathy in subjects with RA. In particular, between 5% and 50% of subjects with RA can experience critical renal failure [127, 128] and injection of gadolinium-based contrast (GBC) agents for subjects with compromised renal systems can be toxic [129]. In another study, RA was diagnosed from MRI scans aimed at detecting

synovitis, bone marrow oedema, and bone erosion. The study reported sensitivity and specificity of 82.5% and 84.5%, respectively [130].

Overall, it has been shown that x-ray, US, and MRI scans are not reliable tools in the diagnosis of RA, and are particularly poor in the diagnosis of early stage RA [83, 125].

3.5 Optical imaging for diagnosis and monitoring

The use of optical techniques for imaging peripheral joints for the purpose of characterizing the presence of RA occurred as a result of studies in the 1990s, when researchers first reported on differences in optical properties between finger joints with and without rheumatic disorders [131]. It was found that in the early stages of RA, changes in the optical properties could be expected in the synovium and the synovial fluid. As already mentioned in Section 3.1, the inflammatory process starts in the synovium, leading to changes in tissue architecture and cell structure. Cell proliferation can be observed. The appearance of the synovial fluid changes from a clear, yellowish substance to a turbid, gray-yellowish substance.

The number of leukocytes, or white blood cells, increases from 100/mL to 200/mL in healthy joints to 1,000/mL to 100,000/mL during stages 1 and 2, respectively. This is particularly important for optical techniques, as leukocytes have a diameter of approximately 7-21 μm and therefore have a considerable effect on the scattering coefficient. Furthermore, the protein content in the synovial fluid approximately triples from 10-20 g/L to 30-60 g/L [132, 133, 134]. In regards to optical imaging of finger joints, the differences in optical signals between healthy joints and joints of subjects affected by RA were strongest at wavelengths between 600 and 700 nm [135, 136].

These indications lead to the development of several optical imaging systems and some pilot studies [131, 137]. In these studies, light from one or more wavelengths were focused onto the back of the finger and transillumination intensity profiles were recorded by a CCD camera. Scheel [138] and Schwaighofer [139] worked on classifying images of finger joints with RA from transillumination images. Scheel [138] and Schwaighofer [139] started to use advanced classification tools

to analyze transillumination images. Using Gaussian process classification, Gaussian process regression, support vector machines, and a generalized linear model they found sensitivities and specificities in the range of 75-85%.

While promising, these studies were limited to monitoring progression of the disease. It appeared that looking for changes could be reliably done, while determining the state of a joint by a single measurement, without a reference (the same joint several weeks before) remained difficult. Schwaighofer states that “the laser images showed high inter-individual variations in optical joint characteristics, resulting from individual differences in joint anatomy.” The transillumination technique seemed ill prepared to deal with these differences.

Some of the limitations of pure transillumination imaging of finger joints were overcome with subsequent DOT imaging studies. Our research team has focused in the past on application of DOT imaging for detecting and characterizing inflammation in RA. Analysis of reconstructed DOT images of absorption and scattering coefficients in the proximal interphalangeal (PIP) joints II-IV has allowed classification of PIP joints with and without RA with sensitivity and specificity of 78% and 76%, respectively [140, 141]. A thorough summary of the research undertaken by our research team will be presented in Section 4.2.

Over the last decade several groups have pursued the use of optical tomographic methods to image arthritis and other joint diseases. Jiang et al. have performed extensive studies to show the potential of optical tomography to detect osteoarthritis (OA). In 2001 they introduced a continuous-wave (CW) system for reconstructing absorption and scattering coefficients of joints [142]. Using experimental data from a human finger and several chicken bones, this group subsequently showed that three-dimensional volumetric reconstructions can provide details of the joint structure and composition that would be impossible from two-dimensional imaging methods [143].

For this study they employed a reconstruction algorithm using the diffusion equation to model for light propagation in tissue. The forward model was solved using the finite-element method (FEM) algorithm. Refining the imaging hardware and software further, they presented initial clinical results involving data from two subjects with OA and three healthy volunteers in 2007 [12, 35].

They found that the DOT images demonstrated differences in optical properties at the joint region between the OA and healthy joints. Since then, this group has further improved their system by introducing system that combines X-ray and optical-imaging [144], developed an instrument for photo-acoustic imaging of joints [145], and moved beyond the diffusion model to include higher-order reconstruction schemes that account for light-transport effects not previously covered [146]. The X-ray system was used to image the distal inter-phalangeal (DIP) joints of 22 subjects with OA and 18 healthy controls, while the photo-acoustic system was use to image DIP joints of 2 subjects with OA and 4 healthy subjects. In 2007, Wang et al. showed the potential of photo-acoustic tomography (PAT) for the imaging of human peripheral joints by studying the method's resolution in cadaver human fingers and small animals [147, 148]. More recently, several researchers have suggested molecular imaging approaches that involve bioluminescence and fluorescence markers [149, 150, 151] as methods to further enhance the ability to diagnose various forms of arthritis. However, these studies did not perform tomographic imaging.

3.6 Discussion

Based on our current understanding of RA and our understanding of the physics behind the interaction of light (photons) and biological tissue, it is our hypothesis that the physiological changes that occur directly as a result of RA are the source of optical contrast that is observed in absorption and scattering coefficient images. This theory motivates our effort to explore the ability to diagnose RA from DOT images. Additional motivation for establishing DOT as a clinically useful tool for diagnosing RA has to do with subject health and comfort: DOT does not expose subjects to ionizing radiation, contrast agents do not need to be used, and imaging is done contact free. As a result, subjects can undergo routine DOT examinations without the risk of adverse side effects and discomfort.

Given the current state of treatment and diagnosis, researchers in the field have called for improved early detection of RA so that disease-modifying anti-rheumatic drug (DMARD) therapies

can be initiated earlier as this could significantly help with the management of the disease [82]. We believe DOT can play a significant role in meeting this need. The work presented in this dissertation is is a first step in achieving this goal.

Computer-Aided Diagnosis

Over the last 25 years computer-aided diagnosis (CAD) has been a subject of intense research for many areas in medical imaging and image analysis [152, 153]. A review of CAD techniques for medical imaging is presented in this chapter. The influence that CAD methods have had on enhancing the clinical values of x-ray CT, US, and MRI images serves as motivation for the application of computational techniques to the analysis of DOT images.

As applied to medical imaging, CAD generally refers to the use of algorithms and software to enhance a physician's ability to detect the presence of disease from medical images. In general, CAD algorithms provide quantitative or qualitative feedback that enhances the ability to accurately diagnose disease. For example, CAD algorithms can simply enhance regions of interest in images for closer inspection by radiologists or other medical personnel, provide quantitative disease scores, or suggest a diagnosis.

Often used to enhance the natural contrast between healthy and diseased tissue, CAD tools have been shown to enhance the diagnostic value of various imaging modalities. Its use in detection and characterization of lesions has been expanded to almost all imaging modalities, including x-ray CT, US, and MRI [154, 155, 156, 157]. CAD was initially applied to radiographs for diagnosis of cardiovascular diseases, lung cancer, and breast cancer [152].

4.1 Traditional Medical Imaging Modalities

The medical applications that have seen the most activity in CAD research are x-ray imaging of the breast, chest, colon, brain, liver, and the skeletal and vascular systems [154]. For example, extensive studies have been performed to evaluate the effectiveness of image processing techniques in increasing the ability of radiologists to detect various pathologies from x-ray images. One application is in the detection of lung nodules from chest radiographs. With the use of CAD, radiologists have been able to detect lung nodules with a 4.6% increase in accuracy (area under the ROC curve increased from 0.894 to 0.940). Alone, the CAD algorithm achieved 80% sensitivity with one false positive per image [154].

CAD has also been applied to the diagnosis of lung nodules from lateral chest radiographs. Diagnosis sensitivities and specificities of 60.7% and 86.9%, respectively, have been achieved [152]. Artificial neural networks (ANN) have been employed in the detection of lung nodules with low-dose computed tomography, yielding sensitivities and specificities of 98.3% and 80.3%, respectively [158]. Application of CAD to the detection of interstitial diseases, detection of interval changes, detection of asymmetric abnormalities, differentiation of lung nodules, and differential diagnosis of interstitial lung disease have also been reported [159].

CAD tools have been particularly successful in enhancing the reading of mammograms. The early detection of breast cancer from mammograms is reported to have been improved by up to 20% when CAD tools are used to aid the diagnosis process [154, 160]. In another study, linear discriminate analysis (LDA) and Bayesian Neural Networks (BNN) were used to investigate the repeatability of CAD based diagnosis of malignant breast lesions with sonography. The best sensitivities and specificities, based on repeatability, were 90% and 66% with BNN, and 74% and 84% with LDA [161], respectively.

Use of CAD algorithms has also enhanced the utility of ultrasound imaging. In CAD enhanced diagnosis of focal liver lesions with contrast-enhanced ultrasonography, classification accuracies between 50.0% and 93.8% were reported [155]. In that study, ANN was used for classification.

CAD tools have also been used to aid in the evaluation of MRI data. Breast MRI (BMRI) has

been the main focus of application for CAD in magnetic resonance. For example, backpropagation neural networks (BPNN) have been used for detection of breast lesions from BMRI scans [156], ANN based CAD algorithms have been used to differentiate between BMRI images of malignant and benign lesions [162], and SVM has been used to study the effect of MRI enhancement thresholding on breast cancer detection rates [163]. Additionally, the ability of CAD to accurately evaluate tumor sizes in breast cancer before and after neoadjuvant chemotherapy has also been explored [164]. Typical sensitivity and specificity values of 73% and 56% have been reported for all cancers [157].

4.2 Optical Imaging

In biomedical optics, CAD has only been applied in a very limited number of studies. For example two papers related to Optical Coherence Tomography (OCT) explored its utility in the diagnosis of esophageal and cervical cancer diagnosis [165,166]. In the first study, a center-symmetric auto correlation scheme was employed to aid in the detection of dysplasia in Barrett's Esophagus disease. Sensitivity and specificity of 82% and 74% were achieved. In another study, the authors extract attributes from three imaging parameters obtained by an NIR imaging system and employ an SVM algorithm to distinguish between malignant and benign lesions [29]. A separate effort has focused on the automated detection of contrast-to-noise ratio regions of interest for DOT imaging of breast tissue [167,168]. In yet another study, the ability to discriminate between breast tissue malignancies using tissue fluorescence and reflectance measurements from diffuse reflectance spectroscopy of excised [169] and in-vivo [170] breast tissue was investigated.

In a separate study, logistic regression (LR) was used in automatic detection of malignant breast lesions by Busch et al. [171]. In that study, 89.0% sensitivity and 94.0% specificity were reported. The average volume of the cancerous tissue was $6.7 \pm 5.2 \text{ cm}^3$, while the average breast volume was $374 \pm 231 \text{ cm}^3$. A total of 35 DOT images of cancer-bearing breast were analyzed. However, in the study, each DOT image of the breast had to be individually segmented in to "cancerous" and

“healthy” regions. Subsequently, each breast image was reduced to two numbers, each representing a region (cancerous or healthy). A leave-one-out cross-validation scheme was used to validate the classifier. Busch recently reported on a follow-up study that explores the utility of a feature called “probability of malignancy,” computed from statistical analysis of total hemoglobin concentration, blood oxygen saturation, and scattering coefficient distributions in female breast from DOT imaging [172]. The study shows that it may be possible to accurately predict the efficacy of neoadjuvant chemotherapy.

To date, our group has been the most active in attempting to introduce CAD techniques to the field of DOT. Over the past six years, our group has studied the use of CAD techniques in the field of DOT, with particular emphasis on the diagnosis of RA. A recent clinical study was performed to explore the feasibility of detecting RA from DOT scans of PIP joints II-IV. In total, 219 joints were imaged and processed. Subsequently, the absorption and scattering coefficient images for each joint were computed and analyzed.¹ The standard for the clinical diagnosis of RA were guidelines set forth by the ACR and the European League Against Rheumatism (EULAR) [10].

Visual inspection of the reconstructed absorption and scattering coefficient distributions did not allow for accurate diagnosis of RA, resulting in low clinical sensitivity (Se) and specificity (Sp) values [10]. This is partly because there is no set criteria for evaluating the presence of RA in DOT images. In general, RA cannot be diagnosed from imaging alone (X-ray, US, or MRI). In practice, a rheumatologist diagnoses RA after multiple consultations and physical evaluation of the compromised finger joints and evaluation of any available radiological imaging (typically X-ray). The inability to recognize the presence of RA from visual inspection has resulted in the need for quantitative techniques that are more capable of evaluating and detecting differences between healthy subjects and subjects with RA.

The difficulties in diagnosing RA from visual inspection of the DOT images alone motivated the development of CAD tools for use in DOT. To date, our research on the use of CAD techniques for diagnosing RA has focused on the classification of constant wavelength (CW) DOT images

¹Full details on the clinical study and subsequent analysis are the focus of Chapters 5, 6, and 7.

of PIP joints [140]. In early work, a small clinical trial was used to obtain data and show that CAD enhanced diagnosis of RA from DOT images might be possible. Basic image features were extracted from absorption images, and linear regression (LR), linear discriminate analysis (LDA), and self-organizing maps (SOM) were used for classification. The results were promising, motivating a larger clinical study to more definitively establish the ability to diagnose RA from DOT images. The initial studies were limited in that only CW-DOT scans were performed on PIP joints. As a result, the utility of scattering images in classification was poor, prompting classification to be performed using only absorption data [140, 173].

The general approach to image classification has been a two-step method: (1) features are extracted from DOT images and (2) features are subsequently used to classify images as belonging to one type of subject or another. Thus far, only four basic image features have been explored: (1) image maximum, (2) image minimum, (3) image variance, and (4) the ratio of maximum to minimum [140].

In previous studies, our research group showed that the absorption coefficients inside the joint cavity might be elevated in subjects with RA compared to healthy subjects [10, 13, 174]. However, using just a single parameter (for example the smallest or the largest absorption coefficient) for classification, sensitivities and specificities of only 71.0% were achieved. Subsequently, Klose et al. showed that if optically derived parameters, such as the minimum and maximum absorption values, are combined for the classification process, sensitivities and specificities can be increased to 76.0% and 78.0%, respectively [140].

Beyond allowing for the diagnosis of RA in well established cases, it is also possible that DOT can aid in the diagnosis of RA in subjects with early symptoms as it is conceivable that changes in the optical properties of the synovial fluid in the compromised joints has started prior to manifestation of the symptoms necessary for the ACR criteria to classify RA. The ACR criteria, on its own, are not sensitive to the early onset of RA (subjects do not yet meet at least four of the seven criteria for classification of RA) [39].

4.3 Discussion

Even with the significant progress reported over the past decade, wide use of DOT for diagnosing and monitoring RA remains elusive and sensitivities remain in the low 80% range. The application of transport theory based reconstruction methods make the image formation and analysis process very time consuming. Both facts remain as major hurdles for wider clinical acceptance. The work in this dissertation addresses these shortcomings in two ways. First, I will develop a CAD toolbox to increase sensitivity and specificity values (Chapters 5, 6, and 7). Second, the image reconstruction process is improved through the development of reconstruction algorithms based on the simplified spherical harmonics (SP_N) equations (Chapters 8, 9, and 10). The SP_N equations are an approximation to the ERT model and can offer substantial improvements in accuracy over the diffusion model in various application.

DOT Imaging Of Joints With RA

The content of this chapter focuses on presenting initial efforts towards achieving Aim 1 of this dissertation, where the objective is to develop CAD tools for DOT that leads to clinically relevant sensitivities and specificities in the diagnosis of RA. This work was published in 2011 on IEEE's Transaction of Medical Imaging [9].

In Chapters 5, 6, and 7 we introduce a general approach to CAD for DOT. We apply this approach to 219 images of finger peripheral interphalangeal (PIP) joints. DOT scans were performed on PIP joints of 20 healthy subjects and 33 subjects with RA [9]. Details of the clinical trial and initial statistical analysis of the data are presented in Chapter 5. The focus of Chapter 6 is to establish a general framework for extracting features of interest from three-dimensional DOT images. Chapter 7 is dedicated to analyzing the ability to accurately classify each joint as affected or not affected with RA through the use of machine learning algorithms, where combinations of multiple individual features from Chapter 6 are used to construct multi-dimensional feature vectors. These vectors are used as input to five different classification algorithms; k nearest neighbors (KNN), linear discriminate analysis (LDA), quadratic discriminate analysis (QDA), support vector machines (SVM), and self-organizing maps (SOM). Algorithm performance is compared in terms of sensitivity and specificity. For each algorithm, we determine the set of features that best differentiates between PIP joints with RA and without RA.

5.1 Background

The initial effort to understand if it was possible to accurately classify DOT images of joints with RA focused on the analysis of data obtained using a continuous wave (CW-DOT or SS-DOT) imaging system. The limitations of CW systems are profound and can result in various difficulties that can limit our ability to diagnose RA. Such difficulties can include an inability to properly separate the absorption and scattering effects. It is well known these limitations can be mitigated by using systems that operate in the time domain (TD-DOT) or frequency domain (FD-DOT).¹

Through collaborations with our colleagues at the University Medicine of Göttingen (Dept. of Nephrology and Rheumatology), Charité University Medicine Berlin (Dept. of Medical Physics and Laser-Medicine), and Georg-August Medical University (Dept. of Radiology), we conducted an extensive clinical study exploring the use of FD-DOT imaging of human finger joints. This study is the first of its kind, as each joint is independently imaged at three different modulation frequencies (0, 300 and 600 MHz).

Like CW-DOT imaging, FD-DOT imaging records the amplitude of the exiting radiation. However, in contrast to CW-DOT, FD-DOT also allows the computation of phase information of the exiting photons. Measurement data is used to reconstruct the three-dimensional absorption and scattering properties of the scanned joint. The reconstruction process relies on using light-propagation techniques that model the propagation of frequency modulated light in tissue. This is particularly important as it has been shown that the additional information provided by phase data improves the separation of scattering and absorption effects [175]. Then, these images are studied for any correlations to the established clinical diagnosis. These results demonstrate that FD-DOT imaging leads to higher sensitivities and specificities than CW-DOT imaging.

The remainder of this chapter is dedicated to presenting details of the clinical study in Section 5.2, the FD-DOT data acquisition and reconstruction methodologies in Section 5.3, and results

¹It should also be acknowledged that there are various applications in which CW-DOT is necessary and advantageous over FD-DOT. Such applications include applications in which high imaging frame-rates are required. This is particularly true when the goal is to monitor physiological processes, such as the rate of change of oxygen concentration levels of blood.

from statistical analysis of the reconstruction data in Section 5.4. The chapter ends with a summary of results in Section 5.5 and a discussion in Section 5.6.

5.2 Clinical study design

5.2.1 Composition of clinical study

The clinical trial included 36 subjects with RA and 20 controls. The study was approved by the Institutional Review Board (IRB) and each participating subject gave informed consent prior to entering the study. Each of the subjects with RA were previously diagnosed with RA and were enrolled at the Department of Nephrology and Rheumatology, University Medicine of Göttingen, Germany. Data from three subjects was discarded because the FD-DOT imaging system failed to operate correctly during the scan.

Of the remaining 33 subjects with RA, 24 were female and nine were male, reflecting the higher prevalence of RA among females compared to males, which is 2 to 3 times more common in women than in men. The mean age of these subjects was 51.5 ± 13.9 years (range 21 to 77 years). All 33 subjects met the criteria for the diagnosis of RA established by the ACR and the EULAR [2, 38]. Most of the subjects with RA (21/33; 63.6%) received DMARD therapy (methotrexate, leflunomide, adalimumab) and low-dose prednisone (< 10 mg/d). Positivity for the RF was found in eight out of 33 cases (24%). All 33 subjects had active disease, defined as at least three swollen and tender peripheral joints and morning stiffness for > 1 hour, with or without an elevated ESR (> 28 mm/hr) or CRP (> 8 mg/l) level. The mean clinical disease activity, assessed according to the method previously defined by van der Heijde et al. in [176], was 4.6 (range 1.59 to 8.02).

All 33 subjects underwent clinical examination (CE), US scans, and low-field MRI imaging of the clinically predominant hand. Additionally, proximal interphalangeal (PIP) joints II-IV were imaged using a sagittal FD-DOT system, resulting in 99 images of fingers from subjects with RA (here referred to as “arthritic” or “affected” joints). PIP joints II-IV can be alternatively referred to by name as the index, middle, and ring fingers, respectively [177]. Fig. 5.1 is a simple anatomical

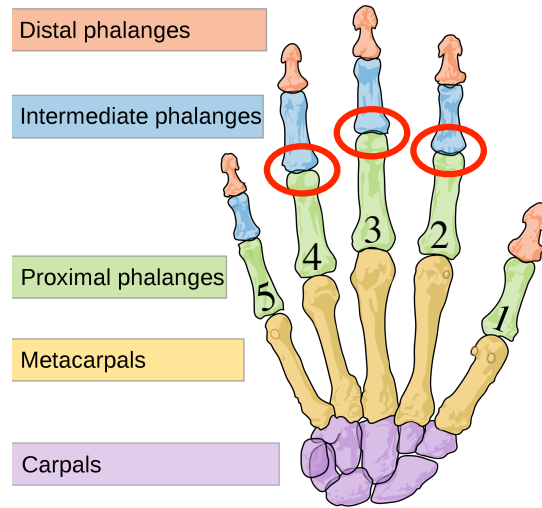


Figure 5.1: Bones of the human hand, highlighting (in red) the PIP joints (II-IV) imaged with FD-DOT. For reference, PIP joint I is the thumb.

atlas of the bones of the human hand provided for reference on the location of the PIP joints relative to the hand.

A control group of 20 healthy subjects was recruited to participate in the study (male to female ratio of 6:14) with a mean age of 38.8 ± 14.1 years (range 22 to 60 years). PIP joints II-IV of both hands of all 20 control subjects were imaged with FD-DOT (resulting in $2 \times 20 \times 3 = 120$ FD-DOT images of healthy PIP joints).

5.2.2 Gold standard

The “true” diagnosis (or gold standard) of each joint is established based on CE, MRI, and US results. The CE of each PIP joint is performed by bi-manual palpation to assess the degree of swelling, tenderness and warmth. The joints are classified according to a clinical synovitis score (CSS) [178, 179]. Laboratory tests (ESR² and CRP³) are used to assess the overall state of the disease. ESR and CRP are both tools for detecting the presence of inflammation (within the body). Higher ESR values indicate the presence of inflammation and occur in inflammatory disease, such as RA. CRP is a protein built in liver tissue and also considered a parameter of inflammation as part

²ESR is defined as the rate at which red blood cells sediment over a period of one hour

³CRP is an acute-phase protein whose concentration levels can be used to monitor the level of inflammation.

of activating the immune system. Although not specific, both parameters show a good correlation to RA disease activity.

Finally, an overall disease activity score (DAS) is assessed as part of the clinical exam. The DAS28 is a combined index that has been developed to measure the disease activity in patients with RA [180, 181]. It includes a classification of 28 joints according to the degree of swelling and tenderness, the ESR, and a patient self-assessment according to the visual analogue scale (VAS). The DAS28 results in a number between 0 and 10, indicating how active the RA is at this moment. Using the DAS28, several thresholds have been developed for high disease activity, low disease activity, or disease inactivity. Disease activity is defined as inactive when $\text{DAS28} < 3.2$, moderately active if $3.2 < \text{DAS28} < 5.1$ and highly active if $\text{DAS28} > 5.1$.

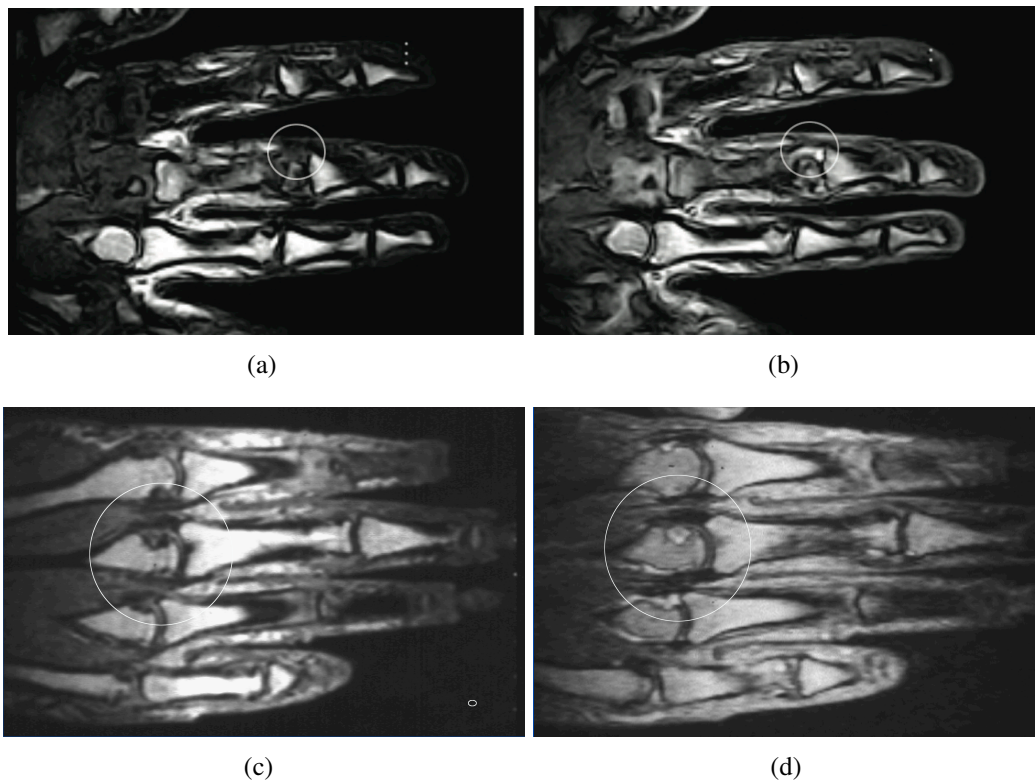


Figure 5.2: (a-b) The two figures show T1-weighted images before (a) and after (b) administration of a bolus of Gd-DTPA. The regions marked with a white circle shows synovitis in a PIP joint. (c-d) The two figures show T1-weighted images before (c) and after (d) administration of a bolus of Gd-DTPA. The region marked with a white circle shows erosions in all three joints inside the circle.

MRI imaging is performed using a dedicated low-field (0.2T) MRI system (C-scan, ESAOTE, Genova, Italy) equipped with a specifically designed hand coil. Imaging sequences included native gradient-echo short-tau inversion-recovery (STIR) sequence in coronal slice orientation, T1-weighted spin-echo high resolution sequence in transversal and coronal orientations, and T1-weighted three-dimensional gradient-echo sequence in coronal slice orientation before and after bolus administration of the paramagnetic contrast medium gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA, Magnevist, Schering, Berlin, Germany) at a dose of 0.2 mmol/kg body weight. The data set acquired with the T1-weighted three-dimensional gradient echo sequence is used for reconstruction of axial views. Examples are shown in Fig. 5.2, which display how synovitis (Figs. 5.2(a), 5.2(b)) and joint erosions (Figs. 5.2(c), 5.2(d)) appears in MR images before and after administration of the contrast agent. Synovitis is clearly apparent in the contrast-enhanced image, while erosion is visible even before the administration of a Gd-DTPA bolus. Scoring of synovitis and erosions is performed according to the EULAR OMERACT criteria [178], using a semi-quantitative scoring system as previously described by Schirmer et al. [179].

US imaging is performed with an Esaote Technos MPX ultrasound system. A 14.0 to 8.0 MHz hockey stick linear array transducer was used for examination of the PIP joints. Examples of US images taken from the palmar side are shown in Fig. 5.3. A healthy joint is shown in Fig. 5.3(a), while Figs. 5.3(b)-5.3(d) show joints with inflammation. In US, two criteria of active inflammation were evaluated following Szkudlarek et al. [116]. Joint effusion (E) was visible as an anechoic area between the capsule and the bone in the proximal part from the palmar side of the hand (Figs. 5.3(b)-5.3(d)). Second, thickening of the synovial membrane (S, synovitis) could be visualized as hyper-echoic structures within the region affected by effusion (Figs. 5.3(d)). US scanning is performed from the palmar side, as it was found that synovitis and effusion can best be evaluated from the palmar as opposed to the dorsal side. This is probably due to the small amount of tissue overlying the joint from the dorsal side.

The US and MRI images were evaluated by a radiologist and a rheumatologist in a blinded-review. The images were evaluated for the presence of effusion, synovitis, and erosion in PIP

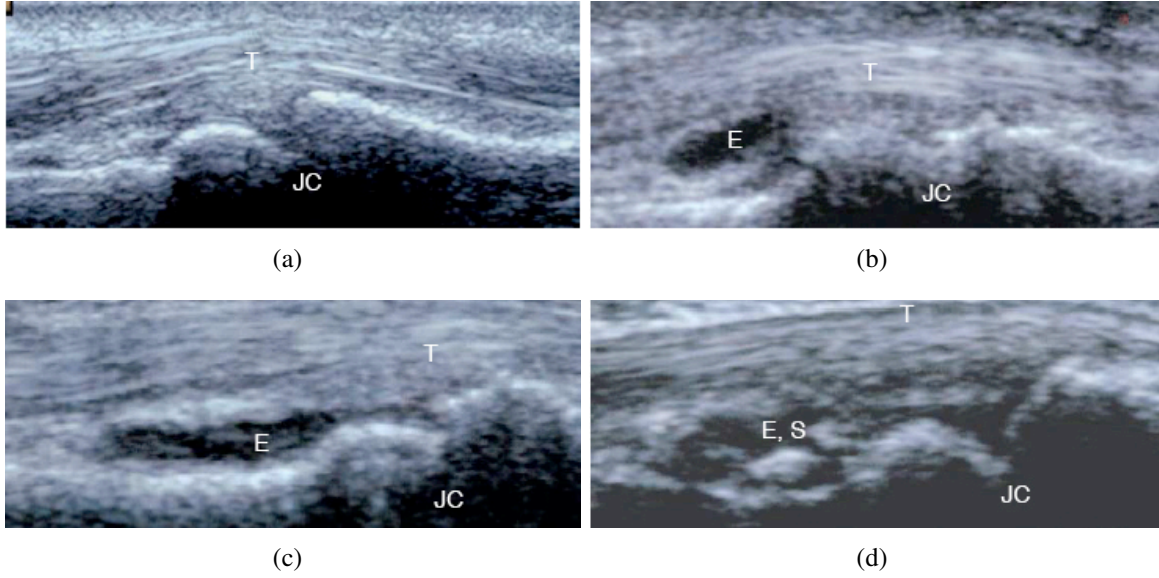


Figure 5.3: Ultrasound images of a healthy joint (a), and joints affected by RA (b,c,d). The larger the anechoic area and/or extent of synovial hypertrophy, the higher was USS (USS = 1 minimal, USS = 2 moderate, USS = 3 extensive effusion/hypertrophy). Images were produced by placing a hockey stick linear array transducer on the palmar side of the PIP joint. (E \equiv effusion, S \equiv synovitis, T \equiv tendon, JC \equiv joint cavity)

joints II-IV. To be classified as affected by RA, joints must show either signs of effusion, synovitis or erosions. Each reviewer classified each subject into one of five sub-groups on the basis of these findings (Table 5.1). The study was designed so that a third reviewer served as a tiebreaker in cases where the initial reviewers had differing opinions (none in this study). Subjects without signs of joint effusion, synovitis, and erosion were divided into two subgroups: (1) subjects with RA and (2) subjects without RA. This classification is used as ground truth for the classification of the corresponding FD-DOT images.

Table 5.1: Six-group diagnostic table based on clinical evaluation and radiological imaging.

Group	Effusion	Erosion	Synovitis	RA	Total Joints
A	No	No	No	Yes	18
B	Yes	No	No	Yes	18
C	No	Yes	No	Yes	12
D	No	No	Yes	Yes	27
E	Yes	Yes	Yes	Yes	24
H	No	No	No	No	120

5.3 Instrumentation and imaging protocol

5.3.1 DOT imaging instrument

DOT imaging is performed with a frequency-domain imaging system that allows for source-modulation frequencies up to 1.0 GHz [182]. A laser beam (wavelength $\lambda = 670$ nm, optical power = 8 mW, beam diameter 1.0 mm) is directed onto the dorsal surface of a finger and scanned across the PIP joint in a sagittal plane (Fig. 5.4). Transmitted light intensities are measured with an intensified CCD (ICCD) camera. The ICCD camera operates in homodyne mode (i.e. the gain of the ICCD is modulated by a slave signal generator at the same frequency as the laser). As a result, a steady state image at the intensifier output is imaged to the CCD. The signal in every pixel depends on the phase between source and detector modulation. Master and slave signal generators are linked together and the phase delay is adjustable. To detect the complete oscillation of the modulation, multiple images are taken at phase delays covering the range of 2π and are transferred to a computer. From the stack of images, two-dimensional amplitude and phase images are derived by data processing. More details concerning this setup can be found in [182].

In addition to scanning the finger for transillumination data, each joint is also scanned by an

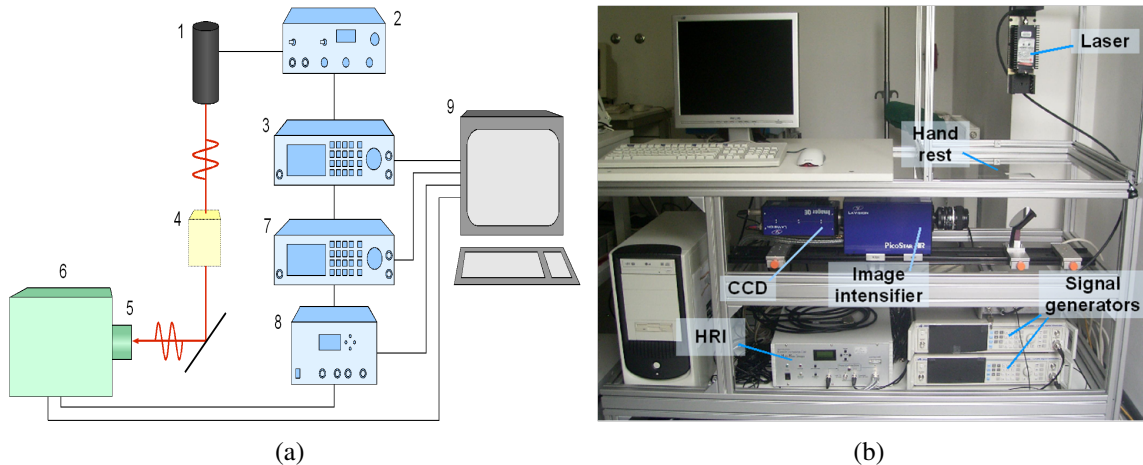


Figure 5.4: Set-up of tomographic scanning unit, schematic (a) and photograph (b). The hardware parts shown are the (1) laser diode, (2) laser diode driver, (3,7) signal generator, (4) finger, (5) focusing lens, (6) ICCD camera, (8) high rate imager, and (9) computer.

independent laser-scanning unit in order to obtain the three-dimensional geometry of the scanned finger (Fig. 5.5a). This is a crucial step in the imaging protocol because the true geometry of the finger is needed for the reconstruction process. The finger is scanned simultaneously by two red laser lines (wavelength $\lambda = 650$ nm, optical power = 5 mW, line width = 0.2 mm) at the dorsal surface of the finger. Each diode laser is mounted on a gear-wheel. A stepping motor generates small rotations of the gear-wheel, resulting in small steps in the laser line on the finger surface. Both gear-wheels are driven by a single stepping motor as this ensures simultaneous scanning of the two lines. The shapes of the deformed laser lines on the finger surface are imaged with a fast video camera (SPC 900 NC, Philips, Netherlands), which is controlled by the DAVID laser-scanner software (version 1.6b, TU-Braunschweig, Germany).

The scanning unit is calibrated (i.e. adjusting the coordinate system of the camera) by using two thin walls on the side of the scanned finger at known angles. The new coordinate system of the camera system is used to transform the observed laser lines into three-dimensional surface coordinates (or point clouds) in real-time.

Scanning of the geometry results in a point cloud that defines the surface points on the finger.

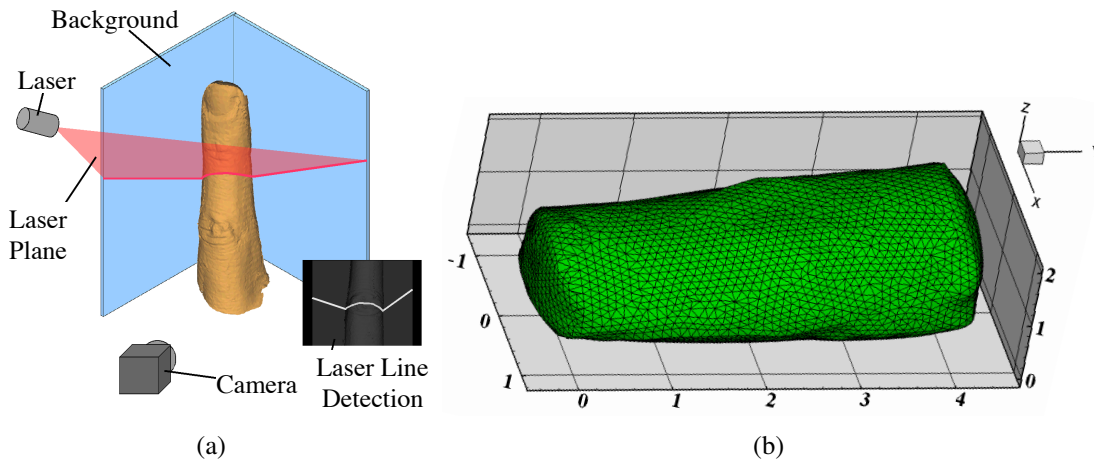


Figure 5.5: (a) Surface registration and (b) three-dimensional mesh generation. The surface scanning unit detects the shape of the laser line on the finger surface and determines the three-dimensional-surface coordinates while the laser line is scanned over the finger. The background serves for calibration of the camera coordinate system. Surface registration and three-dimensional mesh generation; (a) three-dimensional laser scanning to obtain a finger joint geometry and (b) three-dimensional finite volume mesh generated using the laser scanned surface mesh.

From this set of coordinate points, a (closed) surface mesh is constructed that accurately captures the exposed geometry of the scanned finger. Then, the surface mesh is used to generate a volumetric discretization of the surface encapsulated by the surface mesh. We employ the standard tetrahedral discretization technique, which is often referred to as a finite-element method (FEM) mesh or simply an unstructured mesh (Fig. 5.5b). This process is done using the GID software package (Micromechatronics Inc. at <http://www.mmech.com/>). The resulting mesh and transillumination measurements are input to the reconstruction algorithm. Tests on standard surface phantoms show that this set-up can resolve surface details with an accuracy of at least 0.4 mm. Both the tomographic unit and the laser-scanning unit are equipped with identical, ergonomic hand and arm rests at the same height to allow for identical positioning of the patient's arm and hand.

5.3.2 DOT imaging protocol

Before FD-DOT imaging, each finger is marked with a small black dot on the dorsal surface of the finger, 17 mm distal from the PIP joint. This mark is used to position the finger identically in both the FD-DOT imager and in the laser scanner unit. Then, the finger is placed inside the FD-DOT scanner using the hand rest to hold the hand. The laser beam is then moved to the marked position. The finger axis (along the length of the finger) aligns with the scanning plane of the laser. Finally, scanning begins by moving the laser to the first source position, located approximately 10 mm distal from the PIP joint. The laser is then scanned over a range of 20 mm, from proximal to distal end of the finger. The laser stops at 11 distinct and equally spaced positions as shown in Fig. 5.6. Each position represents a unique "source."

Transillumination (on the palmar surface) is recorded for each source position with an intensified CCD camera. For every source position, the oscillation is sampled in 16 phase steps with an exposure time of 80 ms each. The scan is performed twice, first in the forward direction with modulation frequency of 600 MHz, and then in the reverse direction at 300 MHz. Two examples of transillumination obtained from a single surface source, as captured by the ICCD-based detector system, on the posterior (or palmar) surface of the finger are presented in Fig. 5.7.

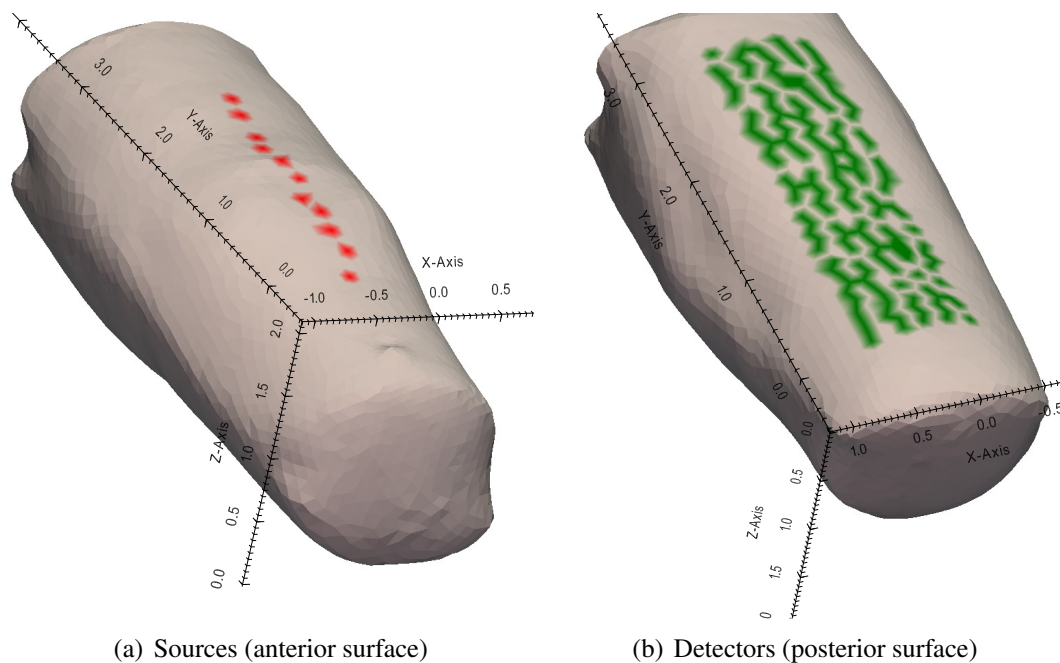


Figure 5.6: Position of 11 sources (a) and detectors (b) on the posterior (dorsal) and anterior (palmar) surface of a sample PIP joint.

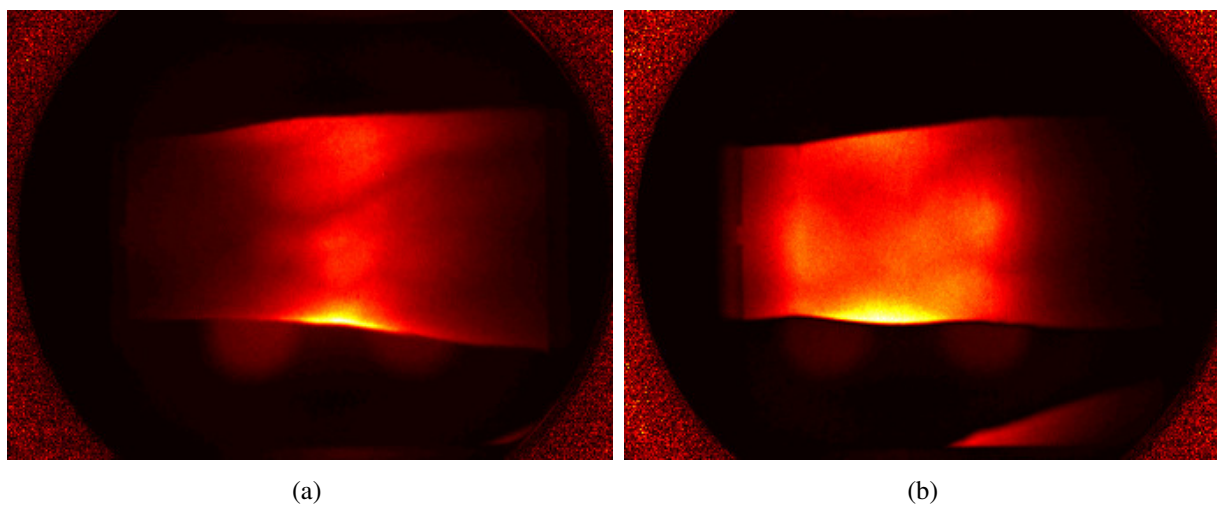


Figure 5.7: Transillumination captured by the ICCD detector unit on the posterior (palmar) surface of a PIP joint from a subject with RA (a) and a healthy control (b).

The positioning of the finger in the laser scanner unit is determined by using the mark on the finger as a reference. An additional laser line, placed across the finger as a pilot beam, helps find the correct axial position for the mark. The finger is angled parallel to the scanning direction. The scan starts approximately 3 mm before the mark and ends after a distance of 40 mm on both sides. One step of the stepping motor yields a step of the laser line on the finger surface of approximately 0.05 mm and takes about 10 ms. Both cameras are in a free running mode and take images with 30 frames per second. A waiting time of 20 ms is inserted between the steps to get approximately one step per frame.

Three fingers were scanned from the predominant hand of each subjects with RA and on both hands of control subjects; the index, middle, and the ring finger (PIP II to PIP IV). To avoid movement artifacts in the image reconstruction, the examiner controls the correct position of the finger again after the tomographic scan is finished. Acquisition time for one finger (including laser movement, image acquisition, and data storage) is about 35 seconds for one frequency. Positioning of the finger averaged another 80 seconds. Thus, the complete tomographic scanning time needed for six fingers and two modulation frequencies was about 15 minutes, or about 7 minutes when using only one frequency.

After acquisition, the raw imaging data is processed. In every stack of images a fast Fourier transformation (FFT) is performed through the stack in every pixel. The FFT yielded values for the amplitude, phase and the DC components. This method allows the computation of two-dimensional DC, amplitude, and phase images for every distinct source position.

5.3.3 DOT image reconstruction

Three-dimensional image reconstruction of absorption and scattering coefficients is performed using an algorithm based on the FD-ERT model. The reconstruction method uses the so-called PDE-constrained reduced Hessian SQP (rSQP) method to solve the forward and inverse problems simultaneously.⁴

⁴ The FD-ERT algorithm is presented in detail by Kim et al. in [20].

A detailed treatment of the numerical techniques for solving the forward and inverse problems in DOT is presented in Chapters 8 and 9. Going beyond a review, novel work is presented on building an algorithm for solving the SP_N light propagation model using the finite volume method and for solving the inverse problem using this same PDE-constrained rSQP method is presented.

However, for completeness, in this section we present a brief overview of the reconstruction algorithm used to obtain the absorption and scattering images shown in this chapter. The PDE-constrained rSQP reconstruction algorithm employs the FD-ERT as the light propagation model. Recalling from Section 2.3.2, the FD-ERT model is given by,

$$(\nabla \cdot \Omega) \psi(\mathbf{r}, \Omega, \omega) + \left(\mu_a + \mu_s + \frac{i\omega}{c} \right) \psi(\mathbf{r}, \Omega, \omega) = \frac{\mu_s}{4\pi} \int_{4\pi} \psi(\mathbf{r}, \Omega', \omega) p(\Omega, \Omega') d\Omega', \quad (5.1)$$

where $\psi(\mathbf{r}, \Omega, \omega)$ is the complex-valued radiance in unit $[\text{W cm}^{-2}\text{sr}^{-1}]$, μ_a and μ_s are the absorption and scattering coefficients, respectively, in units of cm^{-1} ; ω is the external source modulation frequency, c is the speed of light inside the medium, and $p(\Omega, \Omega')$ is the scattering phase function that describes scattering from incoming direction Ω' into scattering direction Ω . The algorithm employs the widely used Henyey-Greenstein phase function with $g = 0.9$ [183].

The model accounts for partially-reflective boundary conditions, allowing for the consideration of a mismatch in refractive index at the interface between tissue and air [61, 184],

$$\begin{aligned} \psi_b(\mathbf{r}_b, \Omega, \omega)|_{\mathbf{n}_b \cdot \Omega < 0} &= \psi^0(\mathbf{r}_b, \Omega, \omega) \\ &+ R(\Omega, \Omega') \cdot \psi_b(\mathbf{r}_b, \Omega', \omega)|_{\mathbf{n}_b \cdot \Omega' > 0}. \end{aligned} \quad (5.2)$$

Where $R(\Omega, \Omega')$ is the reflectivity at a Fresnel interface from direction Ω' to direction Ω , $\psi^0(\mathbf{r}_b, \Omega, \omega)$ is the radiation intensity due to the external source function (i.e. the input source), subscript b denotes the boundary surface of the medium, and \mathbf{n}_b is the unit normal vector pointing outwards at the boundary surface.

Given the spatial distribution of optical properties inside the medium, we solve the ERT (5.1) with a discrete ordinates method [61], which provides the prediction of measurements obtained on

the surface of the medium $P_{d,s}(\mathbf{r}_b, \omega) = Q_d \psi_s(\mathbf{r}_b, \boldsymbol{\Omega}, \omega)$. Here Q_d is the measurement operator that projects the radiance vector $\psi_s(\mathbf{r}_b, \boldsymbol{\Omega}, \omega)$ of a forward model onto the image plane of a CCD camera.

In PDE-constrained optimization theory, the inverse DOT problem requires computation of the radiation intensity vector,

$$\psi = (\psi^1, \psi^2, \dots, \psi^m), \quad (5.3)$$

and the tissue absorption and scattering coefficient vectors,

$$\mu = (\mu_a^1, \mu_a^2, \dots, \mu_a^n, \mu_s^1, \mu_s^2, \dots, \mu_s^n), \quad (5.4)$$

such that

$$\begin{aligned} \min \quad & f(\mu, \psi) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{d=1}^{N_d} (Q_d \psi_s - z_{s,d})(Q_d \psi_s - z_{s,d})^* \\ \text{s.t.} \quad & A(\mu_a, \mu_s) \psi_s = b_s; \quad s = 1, \dots, N_s. \end{aligned} \quad (5.5)$$

Here, (5.5) is the objective function that quantifies the error between the predicted measurement data generated with the current estimate of the inverse variable and the measurement data. N_s denotes the total number of unique boundary sources, and N_d represents the total number of detectors per source. In this work, $N_s = 11$ and N_d are the same for each source (in general, N_d does not need to be the same for each source). In this work we define over 150 mesh nodes as detectors for each source ($N_d > 150$). The terms $z_{s,d}$ and $Q_d \psi_s$ are the true detector measurements and the simulated (i.e. generated with the forward model) predictions for source-detector pair (s, d) , respectively. The operator $()^*$ denotes the complex conjugate of the complex vector inside the parenthesis.

Given the current estimates of forward and inverse variables (μ^k, ψ^k) at inverse iteration step k , the rSQP scheme generates the new iterate $(k + 1)$ for both forward and inverse variables by updating the current iterate:

$$\begin{aligned}\psi^{k+1} &= \psi^k + \alpha^k \Delta\psi^k, \\ \mu^{k+1} &= \mu^k + \alpha^k \Delta\mu^k,\end{aligned}\tag{5.6}$$

where the step length α_k provides a sufficient decrease in the l_1 merit function, and a search direction $\Delta p = (\Delta\psi, \Delta\mu)^T$ can be obtained by solving the quadratic programming problem:

$$\begin{aligned}\min \quad & \Delta p^{kT} g^k + \frac{1}{2} \Delta p^{kT} W^k \Delta p^k \\ \text{s.t.} \quad & C^{kT} \Delta p^k + (A\psi - b)^k = 0.\end{aligned}\tag{5.7}$$

Here g denotes the gradient of f , W^k denotes the full Hessian (or approximations) of the Lagrangian function $L(\psi, \mu, \lambda) = f(\psi, \mu) + \lambda^T (A\psi - b)$, and C represents the matrix of constraint gradients. A detailed description of the rSQP algorithm for ERT-based FD-DOT is presented by Kim et al. in [20].

The algorithm requires as input the three-dimensional unstructured mesh of each finger joint (i.e. an ‘‘FEM’’ mesh), as described in Section 5.3.1 (also Fig. 5.5). The mesh consists of coordinates of each mesh node and connectivity information that specifically indicate which set of nodes (four in the case of tetrahedral elements) form a volume element. A so-called ‘‘finite-volume mesh’’ (FVM) is then generated from the FEM mesh. The FVM mesh is used as the computational grid on which the DOT problem is solved. Detailed information on the FVM method is presented in Chapter 8 in the context of solving the SP_N light propagation model.

The known coordinates of each of the 11 source locations is used to identify the unstructured mesh node closest to that location (on the dorsal surface of the joint). An example of mesh nodes identified as source nodes are shown in Fig. 5.6(a). Then, these 11 mesh nodes are input as source positions to the reconstruction algorithm in the form of $\psi^0(\mathbf{r}_b, \boldsymbol{\Omega}, \omega)$ in (5.2).

Then, the transillumination images captured by the ICCD camera are mapped to the palmar

surface of the unstructured mesh. All mesh points on the palmar side of the finger that map to a pixel on the ICCD image are identified as “detector” nodes. Each of these detector nodes are assigned a “measurement” value corresponding to the data captured by the ICCD image at that location (composed of both amplitude and phase); this information is captured by the term z in the objective function (5.5). An example of mesh points identified as detector nodes are shown in Fig. 5.6(a).

All reconstructions are started with an initial estimate of $\mu_a = 0.3 \text{ cm}^{-1}$ and $\mu_s = (1-g)\mu_s = 8 \text{ cm}^{-1}$ for all mesh points. Typically, the total reconstruction time is typically over 180 minutes on a computer with an Intel Xeon 3.3 GHz processor. A typical three-dimensional unstructured mesh is composed of approximately 30,000 tetrahedral elements.

5.3.4 Diagnostic terminology

Computation of the sensitivity (Se) and specificity (Sp) of a diagnostic scheme is important as these quantities inform on the method’s reliability. These parameters are computed from a “diagnostic table” and rely solely on the number of accurate and inaccurate classifications. In general, each imaged joint has been previously diagnosed as either positive or negative for RA (i.e. the gold standard or ground truth). Then, each DOT image is classified as positive or negative by the classification scheme. If the joint was *positive* for RA, then an image that is classified as “positive” for RA is labeled a true positive (TP), while it is labeled a false negative (FN) if it is classified as “negative” for RA. Similarly, if the joint was *negative* for RA, then an image that is classified as “negative” for RA is labeled a true negative (TN), while it is labeled a false positive (FP) if it is classified as “positive” for RA. Sensitivity and specificity are defined by these quantities, as follows [185].

$$Se = \frac{TP}{TP + FN} \quad (5.8)$$

$$Sp = \frac{TN}{TN + FP} \quad (5.9)$$

Sensitivity is a measure of the diagnostic method's ability to accurately label a positive sample as positive, while specificity is the ability to accurately label a negative sample as negative. Another important quantity is called the Youden index (Y), as it captures the combined effects of sensitivity and specificity on a scale from -1 to $+1$, and is defined as

$$Y = Se + Sp - 1. \quad (5.10)$$

In this formulation, sensitivity and specificity have equal weights and, individually, can range between 0 and $+1$, with 0 corresponding to complete misclassification of all samples and $+1$ corresponding to accurate classification of all samples. Thus, sensitivity and specificity value as close to $+1$ as possible are desirable. Obtaining sensitivity and specificity values of $+1$ is typically difficult, and often, the value of the sensitivity or specificity is increased at the expense of the other.

An important decision in the design of CAD algorithms is the relative importance of sensitivity and specificity. In certain circumstance, such as in the diagnosis of potentially terminal diseases, it may be preferable to have higher specificity at the expense of sensitivity. This scenario may occur because we prefer to err on the side of safety, which in this case means accepting the possibility of more “false positives” in exchange for accurate diagnosis of the disease in additional subjects. In other cases, however, the opposite may be true. Mainly, that the benefits of correctly diagnosing additional subjects with the disease does not outweigh the costs (in a general sense) associated with false positives.

In the case of RA, we have opted for equal weights to the sensitivity and the specificity. Therefore, in general, we seek to maximize the Youden index instead of the sensitivity and specificity

individually. However, we primarily comment on the sensitivity and specificity values as they are more accurately inform us on the performance of the classifier. Exceptions to this rule are explicitly mentioned throughout this work.

5.3.5 ROC curve analysis

Receiver operator characteristic (ROC) curves help understand the diagnostic accuracy of an clinical marker (or image feature in this work). This method allows, in its simplest form, to measure the ability to discriminate between two classes (e.g. healthy versus affected by RA) using a single image feature (i.e. $\min(\mu_a)$ or other similar features) [185]. A characteristic or curve is generated by plotting all of the sensitivity and specificity pairs that result from continuously varying the decision threshold over its entire range (i.e. the value of the feature that separates affected from healthy). The plot itself is the sensitivity (Y-axis) versus $1 - \text{specificity}$ (x -axis) and shows the overlap between the two distributions (i.e. overlap between affected and healthy at the five threshold).

A threshold that perfectly separates the two diagnostic classes results in sensitivity of 1.0 and specificity of 1.0 (or $1 - \text{Sp} = 0.0$), which corresponds to the Cartesian coordinates of $(0.0, 1.0)$ on the ROC plot. An image feature with perfect discrimination for its entire range results in a horizontal line as $y = 1.0 \forall x \in [0.0, 1.0]$. Conversely, a feature that has shown no ability to discriminate between the two classes results in sensitivity and specificity of 0.5 for all threshold values, resulting in a 45° line from $(0.0, 0.0)$ to $(1.0, 1.0)$. Qualitatively, the closer the curve is to the upper left-hand corner (or to the point at $(0.0, 1.0)$), the better the classification accuracy.

For example, assuming a certain threshold $[\min(\mu_a)_{th}]$ for the feature $\min(\mu_a)$, we say that all images with $\min(\mu_a) < \min(\mu_a)_{th}$ belong to healthy subjects, while all other images belong to subjects affected by RA. Then, we calculate the number of TP, TN, FP, and FN using this classification methodology (Fig. 5.8a). Given these numbers, we determine the more clinically significant values for Se and Sp as outlined in Section 5.3.4. By varying the threshold values from 0 to the largest value of $\min(\mu_a)$, we obtain a series of Se and Sp values, which we plot as a ROC

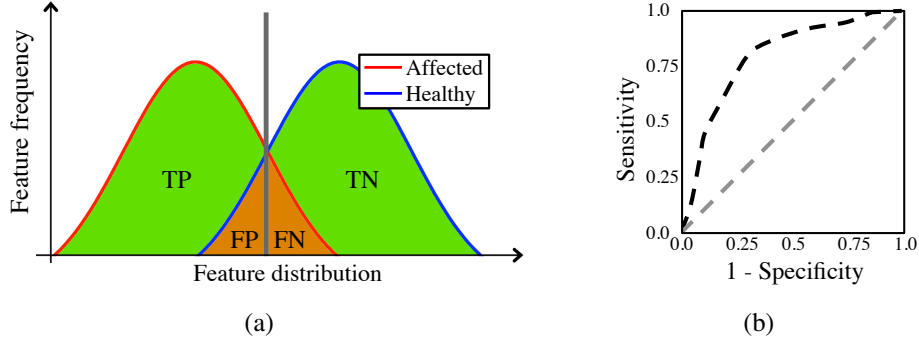


Figure 5.8: (a) Example of how TP, TN, FP, and FN can be computed given a threshold value (vertical line). Se and Sp can be computed from those values. (b) By continuously varying the threshold value, Se and Sp pairs can be computed and used to plot ROC curves. The 45° line is shown for reference.

curve (Fig. 5.8b). The plot shows Se and $(1 - Sp)$ for all threshold values. This curve itself can further be described by the area under the curve (AUC) and the Youden index, which is equivalent to the threshold for which $Y = Se + Sp - 1$ is largest.

5.4 Data analysis

5.4.1 Region of interest and feature extraction

Dimensionality reduction is necessary because it is impossible to compare joints using all of the reconstruction data (i.e. all pixels) for two reasons. First, each mesh is unique to each joint, and as such, comparisons between joints on a pixel-to-pixel basis is not feasible.⁵ Secondly, the number of “features” that would arise if we were to consider each pixel as a feature would be much larger than the number of unique samples (i.e. joints), thus creating a statistical problem where we have significantly more observations than samples.⁶ This would inevitably lead to over-fitting problems.

Instead, we consider only a small number of image features that can be computed from all reconstruction images. After reconstruction of the three-dimensional spatial distribution of μ_a and μ'_s , a three-dimensional region of interest (ROI) is determined for each finger. Only mesh points

⁵This problem could be mitigated by generating a standard joint map and transforming all imaged joints onto that standard map. That strategy, however, remains beyond the scope of this analysis.

⁶This strategy presupposes that all joints are discretized with the exact same mesh.

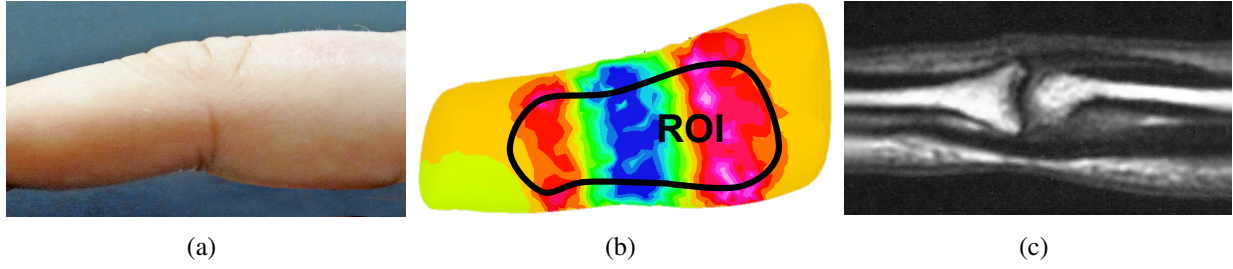


Figure 5.9: (a) Photo showing the approximate size of a finger for which an FVM mesh is generated and used to reconstruct the FD-DOT image. (b) Example of a two-dimensional cross-section through a three-dimensional reconstruction of the absorption coefficients in a finger joint. Also shown is the ROI from which various optical parameters are determined. The ROI is limited to regions at least 2 mm away from boundaries and within the lateral extent of the source and detector location. (c) For further orientation of the imaging range, the corresponding sagittal cross-section from an MRI scan is shown.

that are at least 2 mm away from the tissue boundary are considered. This helps eliminate any reconstruction artifacts that may occur near tissue boundaries. Furthermore, only mesh points that are within the lateral extent of the light illumination points are included in the ROI. Only within this range do the measurements provide useful information for the reconstruction code. Fig. 5.9b is an example of the computed absorption map and the resulting ROI for a finger. For better orientation, we also show a photograph and an MRI image that covers approximately the same area of the joint and finger.

From each three-dimensional ROI we extract four parameters (i.e. “image features”) that are subsequently analyzed. The first two features are the maximum (max) and minimum (min) of the optical property within the ROI. The third feature is the ratio between maximum and minimum ($ratio$). The fourth feature is the variance (var) across all pixels in each three-dimensional ROI volume. These four features are extracted from μ_a and μ'_s images, resulting in a total of eight image features.

Thus, the reconstruction image of each joint is transformed into a vector of eight values corresponding the extracted features (min , max , $ratio$, and var) of absorption and scattering, respectively. Statistical analysis is then performed on these vectors. The general process is depicted by the schematic in Fig. 5.10.

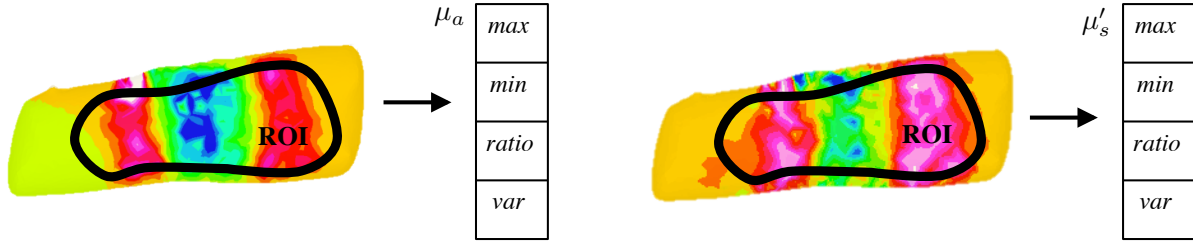


Figure 5.10: Schematic of the feature extraction process. Four features are extracted from the absorption and scattering images of each joint, resulting in a vector of eight image features.

5.4.2 ANOVA and student-t test

Additional analysis is performed by reducing the number of diagnosis subgroups from six to three, by combining subgroups B, C, D, and E into one single group. This simplification was motivated by the physiological differences between group A (subjects that have RA, but whose joints do not have effusion, erosion, or synovitis) and groups B to E (subjects that have RA and whose joints exhibit effusion, erosion, or synovitis). We refer to group A as “asymptomatic” (or simply group A), as these joints have been diagnosed with RA but do not exhibit radiological symptoms. We refer to the group that results from combining groups B to E as “symptomatic” (or simply group S), as these joints belong to individuals with RA and have radiological signs of RA related symptoms. The third group is simply the group of healthy control subjects (or simply group H).

Thus, to be classified as “symptomatic” and in group S, joints have to show signs of effusion, synovitis or erosions. Joints classified as “asymptomatic” and in group A do not show any of these signs. In total, there are 81 fingers in group S, 18 fingers in group A, and 120 fingers in group H. See Table 5.2 for a summary.

We perform ANOVA (imbalanced one-way ANOVA using Tukey’s test in a multiple compari-

Table 5.2: Three-group diagnostic table based on clinical evaluation and radiological imaging.

Group Name	Effusion, Erosion, or Synovitis	RA	Total Joints
A	No	Yes	18
S	Yes	Yes	81
H	No	No	120

son procedure) to understand the differences, if any, between the six diagnostic groups in Table 5.1, and, separately, the three diagnostic subgroups in Table 5.2. We also compute the F-statistic of the data to show which, if any, groups are statistically different. The analysis is repeated for each of the four image features from μ_a and μ'_s images, respectively. In all cases, one feature is considered at a time (i.e. *min*, *max*, *ratio*, or *var*) and we calculate the mean and variance of the feature vectors of each group.

In the case of the three group data, we perform additional analysis by comparing only groups S and H using the student-*t* test, from which we derive *p*-values for each image feature and determine if a statistically significant difference between affected and healthy fingers exists.

5.4.3 Image classification

Clinically, image classification is more relevant than ANOVA and student-*t* test analysis as it informs us of the potential diagnostic value of each image feature, allowing for computation of diagnostic sensitivities and specificities. However, image classification requires that we consider a two-class problem, where one class is “positive” for RA and the other is “negative” for RA. As a result, we only consider groups S and H for this analysis.

We perform ROC curve analysis to study the performance of each image feature as a one-dimensional classifier. Then, we perform LDA classification to understand the performance of multi-dimensional feature combinations as classifiers. Refer to Section 5.3.5 for details on ROC curve analysis and Section 7.2.2 for a summary of the theoretical aspects of LDA.

The use of LDA and multi-dimensional classification is motivated by our previous studies, where we show that it can be advantageous to combine several of the image features as it may result in more significant sensitivity and specificity values [140, 173]. Thus, in addition to comparing Se, Sp, and Y for specific individual image features, we also study the Se, Sp, and Y derived from combining two or more of these individual parameters. To do so we employed two-, three-, and four-dimensional LDA [186, 187].

We use the *leave-p-out* cross-validation technique to avoid over-fitting errors when using LDA.

This process requires that we randomly select $(100 - p)\%$ of the data set to calculate the separation line (or hyperplane) as shown in Fig. 5.17. The remaining $p\%$ of data points are subsequently classified by calculating the TP, FP, TN, FN, Se, Sp, and Y values. We perform this procedure 100 times for each set of features; therefore, 100 times we randomly select $(100 - p)\%$ of the data set to determine the optimal separation line and use the remaining $p\%$ of the data to calculate the TP, FP, TN, FN, Se, Sp, and Y values. We report the performance of the algorithm by the average values for Se, Sp, and Y over all 100 iterations. We use $p = 10\%$ throughout this work

5.4.4 Modulation frequency analysis

As was previously mentioned, all subjects were imaged at 600, 300, and 0 MHz. Thus, our data set consists of reconstruction images of μ_a and μ'_s computed from data at each of the three modulation frequencies. To better understand the differences between the images obtained at each of the three modulation frequencies, we perform the above analysis on the data computed from each frequency.

Then, we compare the ability to diagnose RA from each of the three frequencies. This allows us to establish which of the three frequencies is optimal for discriminating between joints of subjects with RA and healthy joints.

5.5 Results

We start by showing examples of two-dimensional cross-sections through the three-dimensional tomographic reconstructions of optical properties of healthy fingers and fingers affected by RA. Figure 5.11 shows images of μ_a , and Fig. 5.12 shows images of μ'_s . All of these reconstructions were computed using measurement data obtained with a source modulation frequency of 600 MHz.

The most striking pattern in these images is that fingers from healthy subjects (Figs. 5.11(a), 5.12(a)) appear to show a larger variation of μ_a and μ'_s values as compared to images of fingers from subjects affected by RA (Figs. 5.11(b), 5.12(b)). Additionally, images of healthy joints appear to show lower μ_a and μ'_s values than joints affected by RA, especially in the region of the joint cavity,

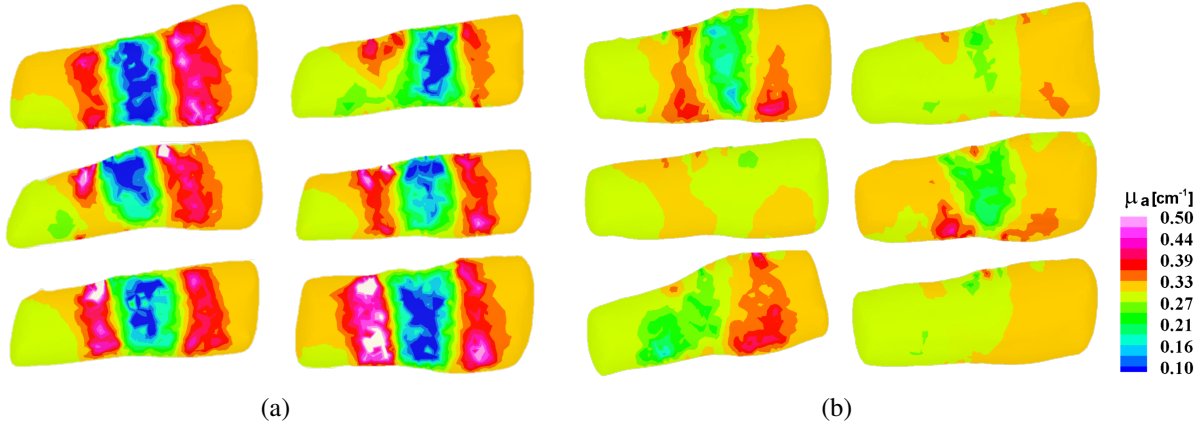


Figure 5.11: Examples of two-dimensional FD-DOT cross-sections through (a) healthy joints and (b) joints affected by RA, showing the spatial distribution of μ_a . Imaging was performed with a source modulation frequency of 600 MHz.

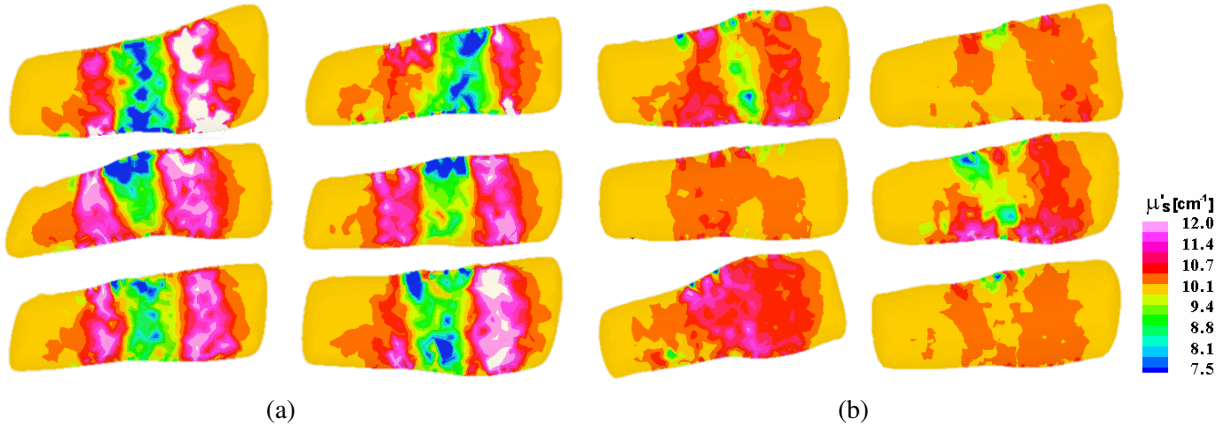


Figure 5.12: Examples of two-dimensional FD-DOT cross-sections through (a) healthy joints and (b) joints affected by RA, showing the spatial distribution of μ'_s . Imaging was performed with a source modulation frequency of 600 MHz.

in the center of the images.

This is in agreement with the well known fact the synovial fluid as well as the synovium of joints in subjects with RA undergo changes in optical properties [135, 136]. The inflammatory process starts in the synovium, leading to changes in tissue architecture and cell structure. Cell proliferation can be observed and the appearance of the synovial fluid changes from a clear, yellowish substance to a turbid, gray-yellowish substance. The number of leukocytes increases from 100/mL to 200/mL in healthy subjects, to 1,000/mL to 100,000/mL during stages 1 and 2 in

subjects with RA. This is particularly important for optical techniques, since leukocytes have a diameter of approximately $7\text{ }\mu\text{m}$ to $20\text{ }\mu\text{m}$ and therefore have a considerable effect on the scattering coefficient. Furthermore, the protein content in the synovial fluid approximately triples from 10 g/L to 20 g/L to 30 g/L to 60 g/L [132, 133]. This has the effect of making the content of the synovial cavity more opaque, which is manifested in these FD-DOT images as an elevation in absorption and scattering.

5.5.1 ANOVA analysis

In this section we present results from ANOVA analysis of the six diagnostic subgroups (A, B, C, D, E, H) presented in Table 5.1 and the three diagnostic subgroups (A, S, H) in Table 5.2, respectively.

Results from ANOVA on all six subgroups is summarized in Fig. 5.13, where the means of each subgroup are plotted along with the 95% confidence interval. Results are provided for μ_a and μ'_s (or simply μ_s) images. Groups whose confidence intervals do not overlap each other's confidence interval are said to differ at statistically significant levels (here at $p < 0.05$). For convenience, the lower and upper bound of the confidence interval of group H is depicted. Thus, group H is statistically different from subgroups whose confidence intervals do not cross those bounds.

Group H is statistically different from groups A, B, C, D, and E in features $\max(\mu_a)$, $\text{var}(\mu_a)$, and $\text{var}(\mu'_s)$. In most cases, group A appears to be in a range between group H and groups B, C, D, and E. However, group A is not statistically different from the individual subgroups at significant levels.

Results from an imbalanced one-way ANOVA (using Tukey's test in a multiple comparison procedure) on the three subgroups presented in Table 5.2 are summarized in Fig. 5.14. The means of each subgroup are plotted along with the 95% confidence interval. Results are provided for μ_a and μ'_s (or simply μ_s) images. Groups whose confidence intervals do not overlap each other's confidence interval are said to differ at statistically significant levels (here at $p < 0.05$). For convenience, the lower and upper bound of the confidence interval of group H is depicted. Thus,

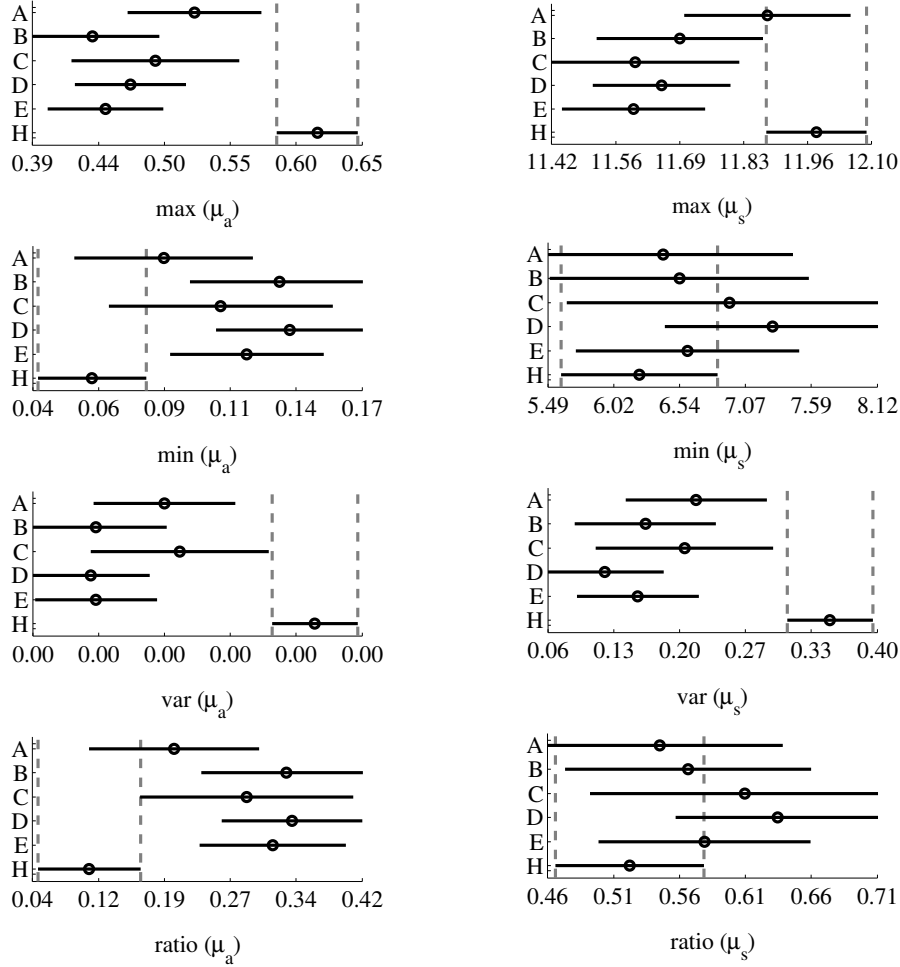


Figure 5.13: Mean values with 95% confidence intervals of all features (600 MHz data) from all six diagnostic subgroups (A, B, C, D, E, and H in Table 5.1). The left column corresponds to features from μ_a images, while the right column corresponds to features from μ_s images. If there is no overlap between groups, then the mean of the given feature is different from the other means at statistically significant levels ($p < 0.05$).

group H is statistically different from subgroups whose confidence intervals do not cross those bounds.

As a reminder, these three subgroups consist of joints from subjects with RA but without radiological symptoms and are therefore treated as “asymptomatic” (A); joints with RA *and* with radiological symptoms - these joints are treated as “symptomatic” (S); and healthy joints (H). Joints in group A are interesting because they appear to be healthy joints when only their MRI and US scans are analyzed. However, these subjects are actually positive for RA.

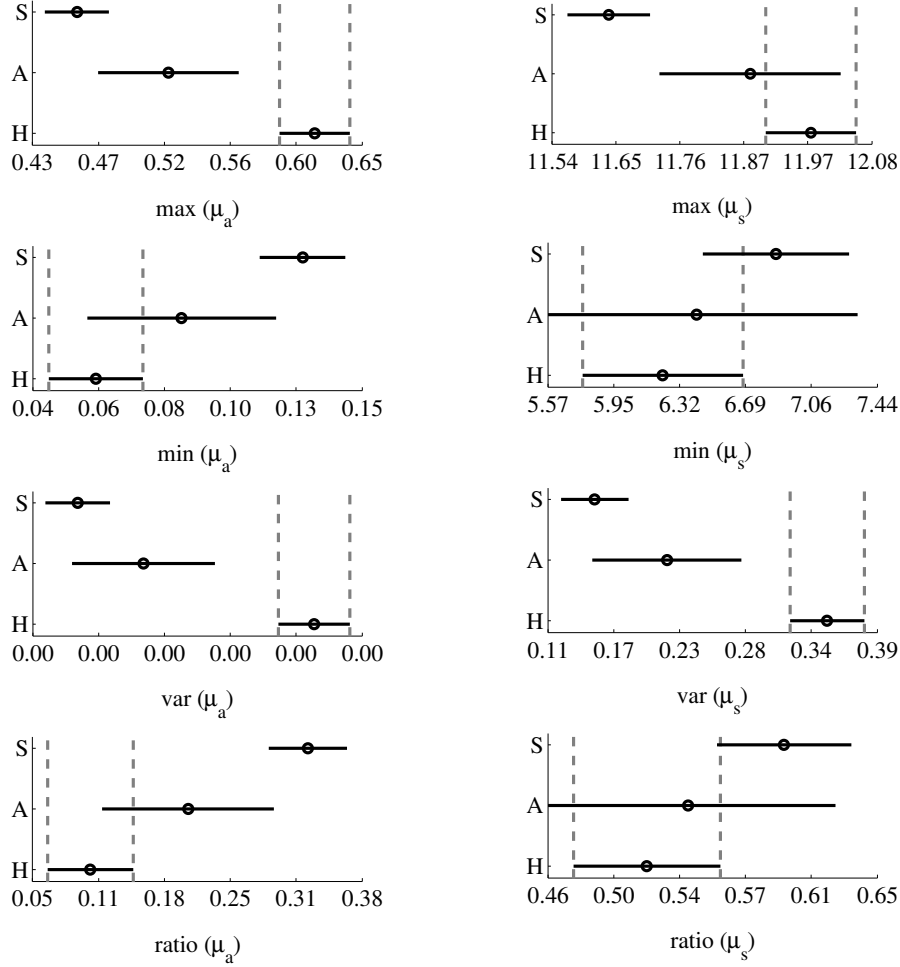


Figure 5.14: Mean values with 95% confidence intervals of all features extracted from FD-DOT images (600 MHz) from groups S, A, and H (see Table 5.2). The left column corresponds to features from μ_a images, while the right column corresponds to features from μ'_s images. If there is no overlap between groups, then the mean of the given feature is different from the other means at statistically significant levels ($p < 0.05$).

We observe that for all features, the mean values of group A lie between the values of group H and group S. In other words, the mean feature value of “asymptomatic” joints is somewhere between the feature values of the “symptomatic” and “healthy” joints. In several cases, the differences are statistically significant. This is particularly interesting because these “asymptomatic” joints appear as “healthy” joints in US and MRI scans, but they clearly do not appear as healthy joints in FD-DOT scans. However, they do not appear as “symptomatic” joints either. Instead, they appear to be a distinct third category of joints.

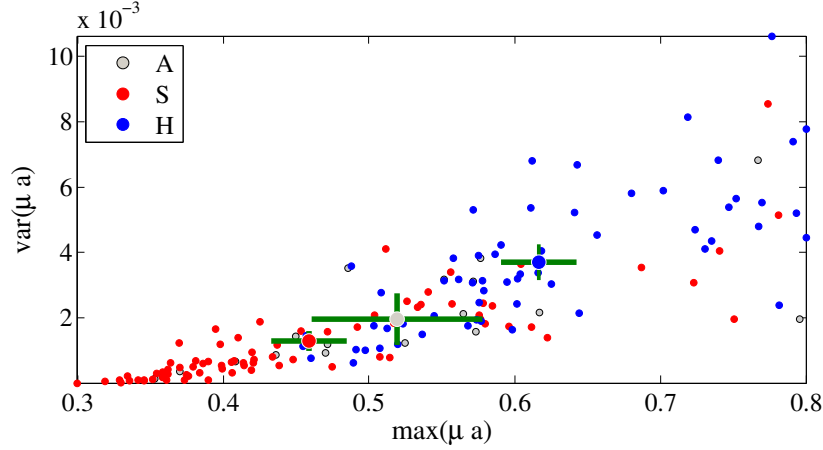


Figure 5.15: Distribution of features found in fingers of healthy (blue dots, group H), symptomatic (red dots, group S), and asymptomatic (gray dots, group A) joints. The bars show the standard error with respect to $var(\mu_a)$ and $max(\mu_a)$, respectively, for each group.

For example, for $var(\mu_a)$, $var(\mu'_s)$, $max(\mu_a)$, the mean of group A differs significantly from the mean of group H; and for $max(\mu'_s)$ there is a significant difference between groups A and S. ROC analysis for this case yields a sensitivity of $Se = 0.79$ and a specificity of $Sp = 0.85$. This suggests that DOT imaging may not only be capable of distinguishing between healthy joints and joints of subjects with RA, but it may also be more sensitive than MRI or US to early changes in joints that do not yet appear to be affected by RA.

Figure 5.15 provides another look at this issue. Here we have plotted the combined image features of $var(\mu_a)$ and $max(\mu_a)$ for fingers in group H (blue dots), group S (red dots), and group A (gray dots). Symptomatic joints show smaller variations and lower maximum μ_a , and hence can be predominately found in the lower left corner of this plot. Healthy joints tend to appear in the upper right corner of this plot, as these joints exhibit larger spatial variations in μ_a and larger maximum μ_a . Data from asymptomatic joints (group A) lie between these two classes. Indeed, looking at the plot we can see that some joints fall clearly within the group of healthy joints, while others fall within the group of affected joints.

This is clinically significant because subjects that do not yet have radiological symptoms may be accurately diagnosed with FD-DOT and recommended to start treatment. Thus, these subjects would begin treatment earlier than otherwise expected if the diagnosis was performed using only

US, MRI, or x-ray imaging. These results are preliminary and studies with a larger number of subjects are need to validate these findings.

5.5.2 Student-t test

Student-t test analysis is used to compare feature vectors of groups S and H (see Table 5.2).⁷ The mean, and standard deviation (*std*) of each feature for each joint is summarized in Table 5.3. We find that, indeed, the differences between healthy and affected joints in $var(\mu_a)$, $var(\mu'_s)$, and $min(\mu_a)$ are statistically significant, with p -values of 2.30×10^{-13} , 2.71×10^{-12} and 2.01×10^{-07} respectively. Interestingly, all features show statistically significant differences ($p < 0.05$) between images of healthy and affected joints, except for $min(\mu'_s)$ ($p = 0.081 > 0.05$).

Table 5.3: Mean and standard deviation of individual features extracted from images of joints from groups S and H.

Absorption Feature Name	Healthy	Affected	p-value
	Group H Mean \pm Std	Group A Mean \pm Std	
Maximum	0.62 ± 0.10	0.46 ± 0.12	3.36E-14
Minimum	0.06 ± 0.05	0.13 ± 0.08	2.01E-07
Variance	0.0037 ± 0.0021	0.0014 ± 0.0013	2.30E-13
Ratio	0.11 ± 0.10	0.32 ± 0.25	3.32E-09

Scattering Feature Name	Healthy	Affected	p-value
	Group H Mean \pm Std	Group A Mean \pm Std	
Maximum	11.84 ± 0.12	11.34 ± 0.51	9.21E-12
Minimum	6.22 ± 1.82	6.86 ± 2.34	8.10E-02
Variance	0.35 ± 0.17	0.15 ± 0.14	2.71E-12
Ratio	0.53 ± 0.16	0.61 ± 0.22	1.20E-02

5.5.3 ROC curve

This section summarizes results from ROC curve analysis comparing features from joints in groups S and H. Figure 5.16 shows the ROC curves for $min(\mu_a)$, $max(\mu_a)$, $ratio(\mu_a)$, and $var(\mu_a)$. The

⁷Here we neglect joints in group U.

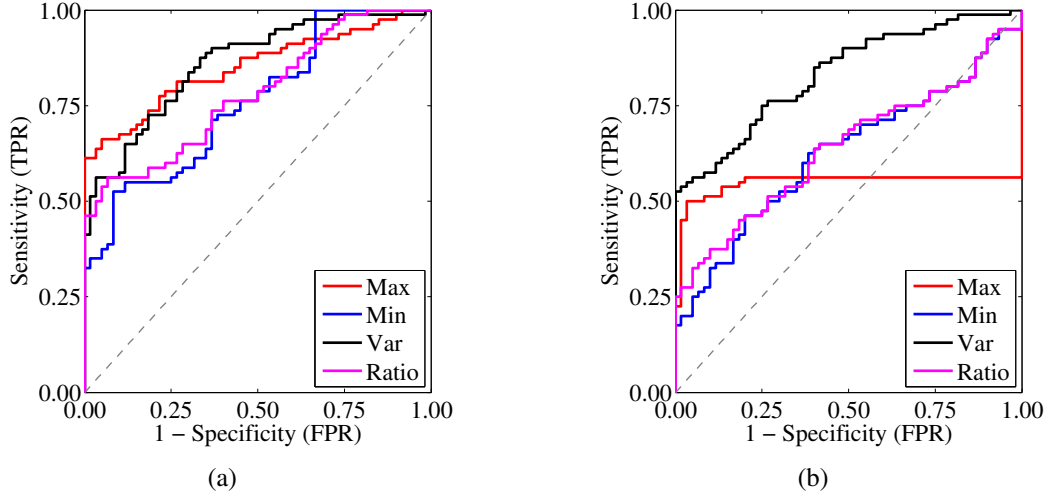


Figure 5.16: ROC curves for the features max , min , var , and $ratio$ extracted from μ_a and μ'_s images obtained at 600 MHz, respectively.

area under the curve (AUC) is largest for $var(\mu_a)$, for which we find $AUC = 0.86$ compared to $AUC = 0.85$ for $max(\mu_a)$. However, $max(\mu_a)$ yields a higher Youden index ($Y = 0.56$) compared to $var(\mu_a)$ ($Y = 0.54$). A similar analysis for μ'_s -dependent features results in lower values for Se, Sp, Y, and AUC . The performance of each feature in ROC analysis is summarized in Table 5.4.

From the Student-t test analysis it is clear that the difference between groups S and H are significant for most features (Table 5.3). However, it is also clear from ROC curves that not all features are equally good classifiers. The two groups are statistically different in all μ_a features. This correlates with all μ_a exhibiting good ROC curve performance as shown in Table 5.4. In contrast, two μ_s features, ($min(\mu'_s)$ and $ratio(\mu'_s)$), show weak or no statistically significant differences,

Table 5.4: ROC analysis results using image features derived from μ_a and μ'_s images (data at 600 MHz).

Absorption					Scattering				
Feature Name	Se	Sp	Y	AUC	Feature Name	Se	Sp	Y	AUC
Maximum	0.78	0.78	0.56	0.85	Maximum	0.56	0.80	0.36	0.55
Minimum	0.55	0.88	0.43	0.76	Minimum	0.63	0.62	0.24	0.63
Variance	0.73	0.82	0.54	0.86	Variance	0.75	0.75	0.50	0.83
Ratio	0.56	0.93	0.50	0.78	Ratio	0.64	0.60	0.24	0.64

correlating with poor ROC curve performance.

5.5.4 LDA classification

Results from classification with LDA and cross-validation using the leave- n -out cross-validation technique are summarized in Table 5.5. This analysis was only applied to data from groups S and H, because the LDA algorithm, as used here, is only capable of acting on a set of data with two labels (i.e. positive or negative for RA). Shown are the Se, Sp and Y values calculated with the LDA method for all possible combinations of μ_a and μ'_s parameters. The numbers in bold text indicate the largest Se, Sp, and Y values observed for both μ_a and μ'_s derived parameters. For example, combining $var(\mu_a)$ with $ratio(\mu_a)$ resulted in $Y = 0.66$ ($Se = 0.83$, $Sp = 0.83$), which is higher than when $var(\mu_a)$ is used on its own, for which we had $Y = 0.58$ ($Se = 0.76$ and $Sp = 0.82$) (Fig. 5.16). When all four parameters are combined a higher specificity of $Sp = 0.86$ is achieved using LDA, however, Se drops to 0.80, yielding the same $Y = 0.66$.

Combining scattering derived image features resulted in even higher Y, Se, and Sp values. For example, we observed that combining $max(\mu'_s)$, $min(\mu'_s)$, and $var(\mu'_s)$ yields the largest Youden index of $Y = 0.76$ as well as the largest Sp value of $Sp = 0.86$. The highest sensitivity $Se = 0.93$

Table 5.5: LDA classification results using image features derived from μ_a and μ'_s distributions obtained with imaging data generated at a modulation frequency of 600 MHz. The numbers in bold indicate the largest values.

Absorption				Scattering			
Feature Combination	Se	Sp	Y	Feature Combination	Se	Sp	Y
max, min	0.79	0.77	0.56	max, min	0.92	0.64	0.56
max, var	0.78	0.79	0.58	max, var	0.76	0.79	0.55
max, ratio	0.8	0.81	0.61	max, ratio	0.93	0.66	0.59
min, var	0.78	0.86	0.64	min, var	0.91	0.82	0.73
min, ratio	0.69	0.67	0.36	min, ratio	0.67	0.69	0.36
var, ratio	0.83	0.83	0.66	var, ratio	0.86	0.74	0.6
max, min, var	0.78	0.86	0.64	max, min, var	0.91	0.86	0.76
max, min, ratio	0.79	0.78	0.58	max, min, ratio	0.92	0.68	0.6
max, var, ratio	0.82	0.81	0.63	max, var, ratio	0.75	0.77	0.52
min, var, ratio	0.8	0.84	0.64	min, var, ratio	0.86	0.82	0.68
max, min, var, ratio	0.8	0.86	0.66	max, min, var, ratio	0.87	0.83	0.7

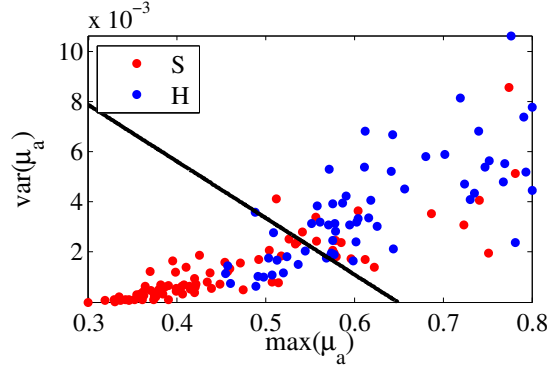


Figure 5.17: Example of LDA decision boundary for discriminating between groups S (red) and H (blue) based on features $var(\mu_a)$ and $max(\mu_a)$.

is reached if $max(\mu'_s)$ is combined with $ratio(\mu'_s)$.

An example of the typical decision boundaries generated by the LDA algorithm is shown in Fig. 5.17, where LDA was used to separate affected joints from healthy joints using the variance and maximum values of μ_a .

5.5.5 Influence of source-modulation frequency

To test the influence of the source modulation frequency on the classification results, we perform the same analysis discussed in Section 5.5.4 with data gathered at 300 and 0 MHz. The results are shown in Tables 5.6 and 5.7, respectively. In general, images from the scans at 300 and 600 MHz yield significantly higher sensitivity, specificity, and Youden index values compared to the images obtained from scans at 0 MHz. This clearly supports the hypothesis that FD-DOT data yields better classification results than CW-DOT.

Comparing the classification results obtained with the 300 and 600 MHz data, we find that overall Se, Sp and Y values are higher when 600 MHz data is used.⁸ This is in agreement with our previous findings, which were based on numerical simulations and phantom experiments, that at source modulation frequencies in the range of 400 to 600 MHz the signal-to-noise level is highest and the resulting image reconstructions are near optimal for geometries encountered in finger

⁸The only exception is the combination of $min(\mu_a)$ and $ratio(\mu_a)$, for which at 300 MHz is $Y = 0.5$, while at 600 MHz it is $Y = 0.36$.

Table 5.6: LDA classification results using image features derived from μ_a and μ'_s distributions obtained with imaging data generated at a modulation frequency of 300 MHz. The numbers in bold indicate the largest values.

Absorption				Scattering			
Feature Combination	Se	Sp	Y	Feature Combination	Se	Sp	Y
max, min	0.78	0.74	0.52	max, min	0.75	0.70	0.45
max, var	0.74	0.76	0.50	max, var	0.75	0.76	0.51
max, ratio	0.74	0.78	0.52	max, ratio	0.58	0.45	0.03
min, var	0.80	0.71	0.51	min, var	0.81	0.74	0.55
min, ratio	0.80	0.70	0.50	min, ratio	0.68	0.63	0.31
var, ratio	0.69	0.77	0.46	var, ratio	0.73	0.77	0.50
max, min, var	0.76	0.72	0.49	max, min, var	0.73	0.73	0.46
max, min, ratio	0.79	0.72	0.51	max, min, ratio	0.75	0.71	0.47
max, var, ratio	0.73	0.82	0.54	max, var, ratio	0.76	0.76	0.52
min, var, ratio	0.79	0.71	0.50	min, var, ratio	0.72	0.69	0.41
max, min, var, ratio	0.78	0.75	0.53	max, min, var, ratio	0.72	0.69	0.42

Table 5.7: LDA classification results using image features derived from μ_a and μ'_s distributions obtained with imaging data generated at a modulation frequency of 0 MHz. The numbers in bold indicated largest values observed.

Absorption				Scattering			
Feature Combination	Se	Sp	Y	Feature Combination	Se	Sp	Y
max, min	0.42	0.33	-0.25	max, min	0.52	0.42	-0.06
max, var	0.55	0.46	0.01	max, var	0.47	0.43	-0.10
max, ratio	0.46	0.42	-0.12	max, ratio	0.29	0.46	-0.25
min, var	0.41	0.37	-0.22	min, var	0.59	0.47	0.06
min, ratio	0.37	0.35	-0.29	min, ratio	0.48	0.44	-0.08
var, ratio	0.43	0.37	-0.20	var, ratio	0.60	0.52	0.12
max, min, var	0.49	0.40	-0.11	max, min, var	0.64	0.51	0.15
max, min, ratio	0.37	0.31	-0.32	max, min, ratio	0.33	0.40	-0.27
max, var, ratio	0.52	0.44	-0.04	max, var, ratio	0.53	0.43	-0.04
min, var, ratio	0.36	0.33	-0.31	min, var, ratio	0.65	0.53	0.18
max, min, var, ratio	0.48	0.39	-0.14	max, min, var, ratio	0.64	0.55	0.19

imaging [77, 188].

5.6 Discussion

In this chapter we presented analysis of clinical data obtained using a new FD-DOT imaging system. In total, 56 volunteers were imaged using a FD-DOT, including 36 subjects with various stages of RA and 20 healthy control subjects [9]. Each subject was evaluated by a rheumatologist

and diagnosed for RA according to guidelines set by the ACR [2]. In the study, the clinically dominant hand of each subject with RA was imaged with US and low-field MRI.

The US and MRI images were evaluated by a radiologist and a rheumatologist in a blinded-review. The images were evaluated for the presence of effusion, synovitis, and erosion in PIP joints II-IV. Each reviewer classified each subject into one of five sub-groups on the basis of these findings (Table 5.1). Each of the 219 joints were assigned to one of six subgroups based on evidence of effusion, erosion, or synovitis in US and MRI images. Then, subjects with RA but without radiological symptoms were assigned to a group of “asymptomatic” subjects (group A), while those with joints with RA and with symptoms were assigned to a group of “symptomatic” joints (group S). A third group of “healthy” joints was created from joints of subjects without RA (group H).

Imaging with a FD-DOT sagittal laser scanner of PIP joints II-IV was performed on the clinically dominant hand of subjects with RA and on both hands of the control group. A frequency-modulated laser beam scanned the dorsal side of the finger from proximal to distal end, stopping at 11 discrete locations to allow for data acquisition. Transillumination was recorded from each source position on the ventral side of the finger with an intensified CCD camera. The three-dimensional geometry of the scanned finger was obtained with a separate laser-scanning unit (650 nm, 5 mW, 0.2 mm line width). Imaging was performed at 0, 300, and 600 MHz. In total, 228 fingers were imaged. Transillumination measurements were used to reconstruct tissue μ_a and μ'_s coefficients. The system and imaging procedures are described in detail by Hielscher et al. [9]. Data from three subjects with RA was discarded due to problems that occurred during the FD-DOT data acquisition process that rendered the measurements unusable for reconstruction. Thus, a total of 219 joints were considered in the analysis.

For each of these joints, transillumination measurements were used to reconstruct three-dimensional absorption and scattering coefficient images using an algorithm based on the ERT light propagation model for NIR light in tissue. The algorithm leverages a PDE-constrained reduced-spaced quadratic programming scheme to expedite the computational process. From each image we ex-

tracted four image features (min , max , $ratio$, var).

In general we found that images generated with 600 MHz data allow for more accurate classification of each joint (i.e. higher Se , Sp , and Y) than images generated with continuous wave data (0 MHz) or 300 MHz. This is in agreement with our previous numerical studies that identified modulation frequencies in the range of 400 to 600 MHz as optimal for finger joint geometries. Those studies, however, relied on numerical simulations and did not include clinical data.

Furthermore, we found that combining several image features results in higher Se , Sp and Y values as compared to using only a single feature. For example, combining $max(\mu'_s)$, $min(\mu'_s)$, and $var(\mu'_s)$ leads to $Y = 0.76$ with $Se = 0.91$ and $Sp = 0.86$. The highest sensitivity $Se = 0.93$ is reached if $max(\mu'_s)$ is combined with $ratio(\mu'_s)$. In general, it appears that scattering-derived features perform slightly better than absorption-derived features in these multi-dimensional feature classifications. This is in contrast to previous results, when only CW-DOT data was considered, where absorption images were determined to provide the strongest classifiers.

Perhaps most importantly, we observed that the image features of joints from group A lie somewhere between the optical properties of joints in group S and group H.⁹ This suggests that optical methods may be useful in diagnosing very early signs of RA in these joints. Larger clinical trials will be necessary to conclusively support this hypothesis.

These results motivate further research into the use of FD-DOT as a tool for the diagnosis of RA. In particular, the use of more sophisticated image features and more advanced machine learning techniques may result in improved sensitivity and specificity values. For this reason, Chapter 6 is dedicated to extending the analysis of the data presented in this chapter by considering features that go beyond the four basic features. The content of Chapter 7 focuses on the use of machine learning techniques that go beyond LDA.

⁹As a reminder, joints in group A are considered “asymptomatic”, meaning that they belong to subjects with RA but do not show signs of effusion, erosion, or synovitis on US and MRI images. Joints in group S are from subjects with RA that do show signs of effusion, erosion, or synovitis on US and MRI images. Joints in group H are from healthy control subjects.

Feature Extraction

Results from Chapter 5 show that FD-DOT images allow for more accurate classification of PIP joints as affected or not affected by RA than CW-DOT images. Furthermore, FD-DOT scans at 600 MHz yield more accurate image classification than data at 300 MHz. While only basic statistical features and a basic multidimensional classification scheme (LDA) are considered, we show that 91.0% sensitivity and 86.0% specificity can be achieved with 600 MHz data. In contrast, the best classification results with CW-DOT data is 64.0% sensitivity and 55.0% specificity.

Furthermore, the study showed that features derived from μ'_s images allowed more accurate classification (91% Se and 86% Sp) when compared to μ_a derived features (83% Se and 83% Sp) [9]. A limitation in that study was that classification was not performed using a mixture of μ_a and μ'_s derived features.

A natural progression of this research is to consider more advanced image features and more sophisticated classification algorithms. We also consider combining features from μ_a and μ'_s data, as up to now, a mixture of these features has not been considered until now.

In this chapter we consider features extracted from the entire three-dimensional μ_a and μ'_s images as well as from projections of these three-dimensional volumes onto two-dimensional planes. We extract three types of features from these data sets. One set of image features is obtained from coefficients of Gaussian mixture models (GMM) that fit the spatial distribution of the optical properties inside the finger. A second set of image features are spatial frequency parameters obtained

through the application of the Fourier transform to these images. Finally, basic statistical features similar to those discussed in Chapter 5 are obtained from all data sets. Overall, a total of 297 features are extracted from each μ_a and μ'_s image (or 594 from each finger).

The statistical significance of each feature is evaluated with classical statistical methods, including Kruskal-Wallis ANOVA, Dunn's test, and ROC analysis. The intra-class correlation coefficient (ICC) is used to compute the effective sample size (ESS) of our data, which in turn is used to adjust our results for any bias that may be introduced by treating each imaged finger as an independent sample. This step is necessary as we imaged multiple fingers per subject. Through this analysis we establish which individual features are best indicators of RA in terms of diagnostic sensitivity and specificity. Links between these best features and physiological processes are identified.

6.1 Clinical data

This analysis is applied to the clinical data set presented in Chapter 5, which includes 219 FD-DOT images of PIP joints from 56 subjects (33 with RA and 20 without *RA*). All subjects are assigned to one of six diagnostic groups (Table 6.1). All subjects are either positive or negative for RA. Subjects with RA are categorized into one of five subgroups based on evidence of effusion, erosion, or synovitis in PIP joints II-IV. Volumetric reconstructions of μ_a and μ'_s within a given finger are obtained for each imaged joint using an ERT-based algorithm (Fig. 6.1).

Examples of cross-sections are shown in Fig. 6.2. The most pronounced differences between

Table 6.1: Diagnostic table based on clinical evaluation and radiological imaging.

Group	Effusion	Erosion	Synovitis	RA	Total Joints
A	No	No	No	Yes	18
B	Yes	No	No	Yes	18
C	No	Yes	No	Yes	12
D	No	No	Yes	Yes	27
E	Yes	Yes	Yes	Yes	24
H	No	No	No	No	120

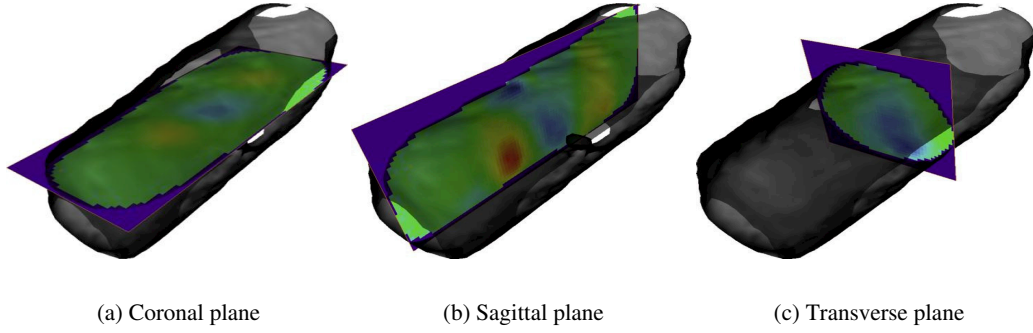


Figure 6.1: Visualization of typical finger geometries. Sample μ'_s cross-sections are visualized within the finger geometry: (a) coronal (xy), (b) sagittal (xz), and (c) transverse (yz) planes.

joints of subjects affected by RA and of subjects not affected by RA occur at the center of the images, the region where the joint cavity is located. As expected, in healthy joints both μ_a and μ'_s often appear to be lower in this region than in the surrounding tissues. The synovial fluid that fills the joint cavity is almost free of scattering and has lower absorption than surrounding tissue. Joints affected by RA typically do not show a drop in optical properties in these regions. We find, however, that relying on visual inspection of DOT images alone does not yield high sensitivities and specificities. The promise of improved clinical diagnostic values motivates research into the application of more advanced CAD techniques.

6.2 Data pre-processing

The reconstructed optical property maps shown in Fig. 6.2 are originally recovered on an unstructured mesh with tetrahedral elements. To simplify our data analysis, first, in a pre-processing step we use interpolation to convert the reconstruction data from the unstructured mesh to a structured Cartesian grid. This is a three step process: (1) define a structured grid that overlaps the tetrahedral mesh; (2) identify the set of structured grid points n_p whose x - y - z coordinates are within the tetrahedral element defined by the set of four unstructured mesh nodes $p = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$, where \mathbf{r}_i refers to the x - y - z coordinates of a node in the unstructured mesh; (3) compute the μ_a and μ'_s values at structured grid point k ($\forall k \in n_p$) using the values of μ_a and μ'_s at each node \mathbf{r}_i of set p .

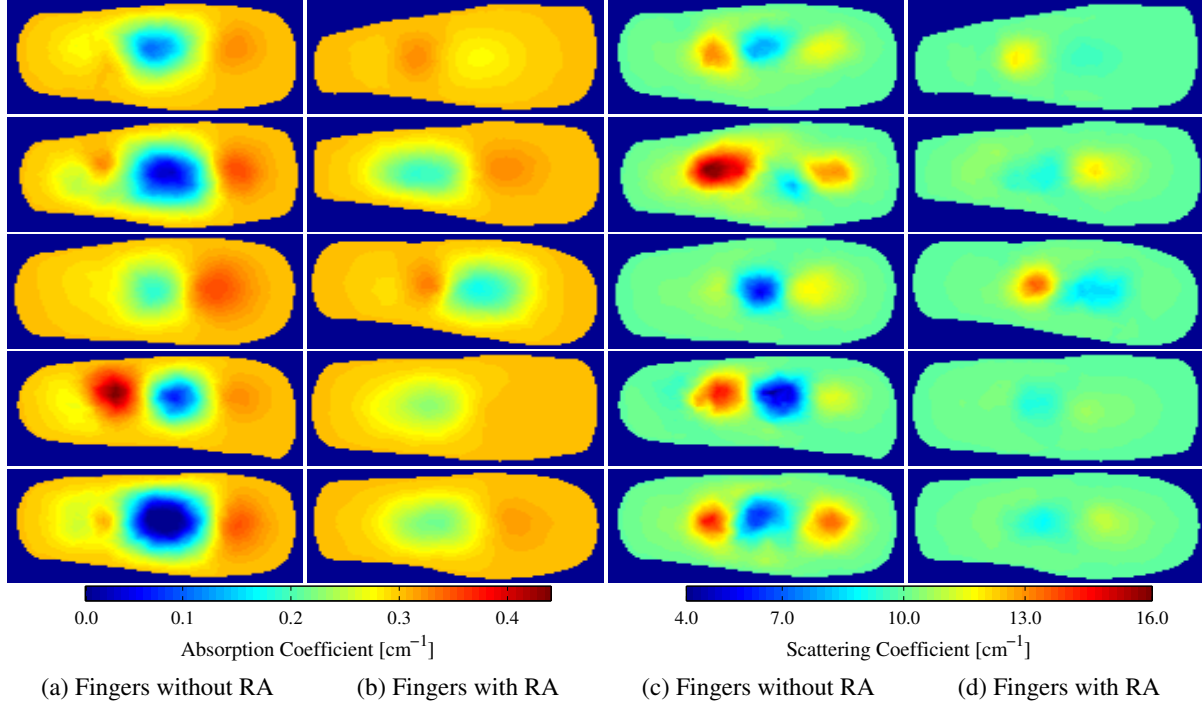


Figure 6.2: (a-b) Absorption and (c-d) scattering coronal cross-sections of PIP joints from subjects without (a,c) and with RA (b,d). All images are reconstructed from data obtained with 600 MHz source modulation.

On the structured grid one can easily define stacks of sagittal (perpendicular to the xz -plane), coronal (perpendicular to the xy -plane), and transverse slices (perpendicular to the yz -plane) as seen in Fig. 6.1. Consider the following example for the rest of this section: structured image A has dimensions $M \times N \times P$ (i.e. number of voxels per axis). There are P coronal slices, N sagittal slices, and M transverse (Fig. 6.3) slices. These slices are “stacks” of images. We apply three pre-processing procedures to each stack.

First we calculate the sum of all sagittal, coronal, and transverse slices, respectively, resulting in three new data sets, which we call SS , SC , and ST (Eq. 6.1). The summation of these slices magnifies regions with large optical parameters inside the finger, as seen in the example in Fig. 6.3.

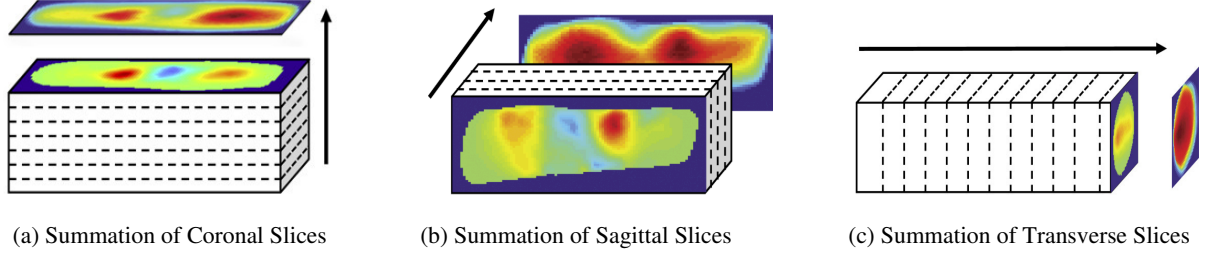


Figure 6.3: An example of the summation of coronal, transverse, and sagittal slices of the three-dimensional data set to create new data sets (a) SC , (b) SS , and (c) ST .

$$\begin{aligned}
 SS_{mp} &= \sum_{n=1}^N A_{mnp} \\
 SC_{mn} &= \sum_{p=1}^P A_{mnp} \\
 ST_{np} &= \sum_{m=1}^M A_{mnp}
 \end{aligned} \tag{6.1}$$

Here, SS_{mp} denotes pixel mp in SS , where $1 \leq m \leq M$ and $1 \leq p \leq P$ ensure that all pixels in SS are defined. The same logic can be applied to interpret SC_{mn} and ST_{np} . Next, we compute the variance between all sagittal, coronal, and transverse slices, respectively. This results in three more data sets called VS , VC , and VT (Eq. 6.2). These data sets quantify the variation between slices, which is a measure of variation in optical parameters inside the finger.

$$\begin{aligned}
 VS_{mp} &= \sum_{n=1}^N \frac{(I_{mp}^{SS} - A_{mnp})^2}{N - 1} \\
 VC_{mn} &= \sum_{p=1}^P \frac{(I_{mn}^{SC} - A_{mnp})^2}{P - 1} \\
 VT_{np} &= \sum_{m=1}^M \frac{(I_{np}^{ST} - A_{mnp})^2}{M - 1}
 \end{aligned} \tag{6.2}$$

As in our previous example, VS_{mp} denotes pixel mp in VS , where $1 \leq m \leq M$ and $1 \leq p \leq P$

ensure that all pixels in image VS are well defined. Furthermore, I^{SS} is the average sagittal slice, where I_{mp}^{SS} denotes pixel mp in I^{SS} and is defined as $\frac{SS_{mp}}{N}$. The remaining variables (VC_{mn} , VT_{np} , I_{mn}^{SC} , and I_{np}^{ST}) are defined in a similar manner.

Finally, data sets GS , GC , and GT are obtained by computing the average of all sagittal, coronal, and transverse slices within ± 2.0 mm from the center of the PIP joint, respectively. In this region, where one typically finds the joint cavity, the differences between subjects affected by RA and healthy volunteers are expected to be the largest. Furthermore, potential artifacts introduced by boundary effects are minimal [19, 65]. The center of the joint is at the geometrical center of the imaged finger, whose dimensions we know from the imaging procedure where the geometry of each finger is captured. Subsequently, we use this geometry to generate the FVM on which we compute the forward and inverse DOT problems [9].

Overall, including the entire volume of the unstructured and structured data, the pre-processing procedures result in 11 distinct data sets per finger (and for each optical variable). The nomenclature used for referencing each processed data set (SV, SS, SC, ST, VS, VC, VT, GS, GC, and GT) is as follows: the first letter indicates the type of pre-processing (S = sum, V = variance, and G = geometrically dominant) and the second letter refers to the physiological plane of the resulting data set (S = sagittal, C = coronal, and T = transverse). Table 6.2 summarizes nomenclature used in this paper and Fig. 6.4 summarizes the 11 distinct data pre-processing steps.

Table 6.2: Summary of data pre-processing nomenclature.

Name	Description	Name	Description
UV	Entire volume (unstructured data)	VC	Variance between coronal slices
SV	Entire volume (structured data)	VT	Variance between transverse slices
SS	Summation of all sagittal slices	GS	Geometrically dominant sagittal slice
SC	Summation of all coronal slices	GC	Geometrically dominant coronal slice
ST	Summation of all transverse slices	GT	Geometrically dominant transverse slice
VS	Variance between sagittal slices		

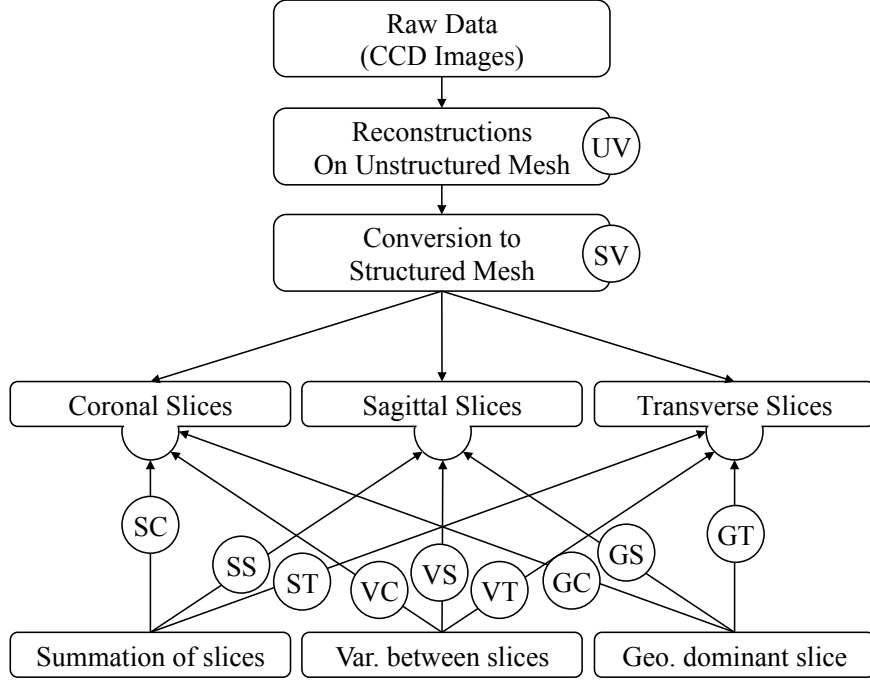


Figure 6.4: Data processing steps, starting with the raw CCD data, followed by processing of the unstructured and structured data sets, and ending with application of the 3 projection operators to the three-dimensional structured data. This results in 11 distinct data sets (represented by each circle and summarized in Table 6.2). The feature extraction operators are applied to each set.

6.3 Feature extraction

We extract three different types of features from all the data sets described in the previous section. These features include: (1) basic statistical values, (2) Gaussian mixture model (GMM) parameters, and (3) Fast-Fourier Transform (FFT) coefficients. Each type of feature is described in more detail in the following sections.

6.3.1 Basic features

The basic statistical features are the maximum, minimum, mean, variance, and the ratio of maximum to minimum of each data set. These features are summarized in Table 6.3, where each feature is assigned a number (#) that is used for referencing throughout this paper. These five features are obtained from each of the 11 data sets (Table 6.2) by arranging the optical parameter into vectors of ascending value. Each reconstructed property, \mathbf{x} , is expressed as $\mathbf{x} = [x_1, x_2, x_3, \dots, x_N]$, where

Table 6.3: Definition of basic statistical features.

#	Feature Name
1	Maximum
2	Minimum
3	Mean
4	Variance
5	Maximum/Minimum

the computational domain has N mesh points and x_i is the optical property at the i^{th} mesh node.

To avoid singular outliers, we calculate the average of the 10 largest and 10 smallest values and assign them as the maximum and minimum features, respectively. The mean and variance are computed from data that does not include the 10 largest and 10 smallest values. The ratio between maximum and minimum is computed as the fifth basic feature.

6.3.2 GMM coefficients

An additional seven features are extracted from all data sets (except the unstructured data) by parameterizing the images with a two-dimensional or three-dimensional multivariate Gaussian mixture model (GMM). Parameterization with GMMs is chosen because the reconstructed distributions of the optical properties are typically smooth varying functions in space. We fit the GMM by finding estimates for amplitude A_0 , covariance matrix Σ , and mean \mathbf{x}_0 of the Gaussian function (G),

$$G(\mathbf{x}) = A_0 \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \Sigma^{-1} (\mathbf{x} - \mathbf{x}_0) \right]. \quad (6.3)$$

Parameters A_0 , Σ and \mathbf{x}_0 are estimated using the Expectation-Maximization (EM) algorithm [189]. The model data allows for more advanced statistical analysis as the entire image is described by only a few parameters (Fig. 6.5). We set the total number of Gaussian functions in the GMM model to 8, as we find that they provide sufficient accuracy.

Table 6.4: Definition of GMM features. (* These feature are only applicable when considering three-dimensional images.)

#	Description
6	Absolute error between original image and GMM image
7	1 st eigenvalue of Σ of largest positive Gaussian
8	2 nd eigenvalue of Σ of largest positive Gaussian
9	3 rd eigenvalue of Σ of largest positive Gaussian*
10	1 st eigenvalue of Σ of largest negative Gaussian
11	2 nd eigenvalue of Σ of largest negative Gaussian
12	3 rd eigenvalue of Σ of largest negative Gaussian*

In the case where we consider data on unstructured grids, it is necessary to modify the algorithm so that each data “point” has a corresponding weight. Using the volume of the FVM node around each node as the weight works well.

Features that described the parameterization of the concave (positive) and convex (negative) regions are extracted, including the absolute error between the mixture model and the original data (Table 6.4). The eigenvalues of the dominant positive and negative Gaussians are computed and extracted, as these features can quantify the spread of the Gaussian functions.

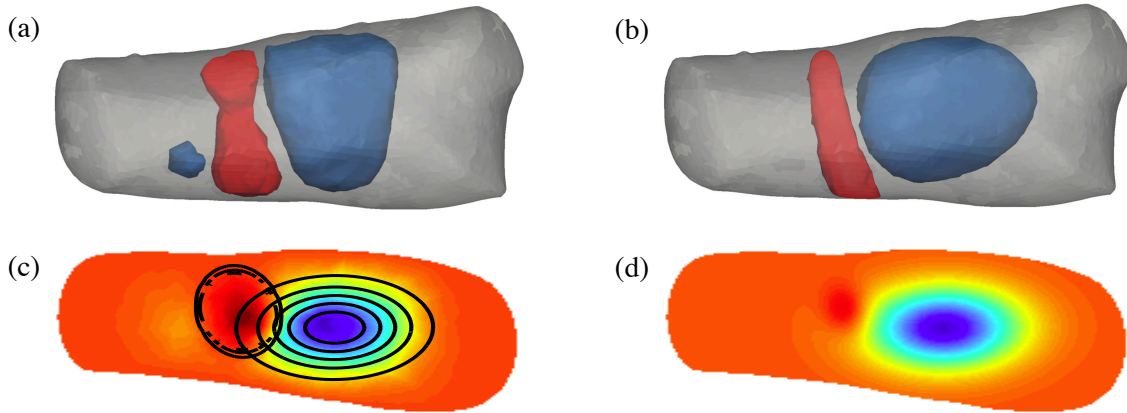


Figure 6.5: (a-b) Three-dimensional example of μ_a around a joint. (c-d) Coronal cross-section of μ_a across the same PIP joint. (a) Iso-surfaces of the original three-dimensional data. (b) Iso-surfaces of the GMM model showing a good approximation to the original data. (c) Isolines are superimposed on the original image to show the resulting fit from GMMs. (d) The model image generated from the coefficients of the GMM model.

6.3.3 FFT coefficients

In addition to using GMMs, a two-dimensional image or three-dimensional volume can also be parameterized by performing a two- or three-dimensional discrete fast Fourier transform (FFT).

The n -dimensional Fourier transform,

$$\hat{f}(\mathbf{k}) = [\mathcal{F}_{x \rightarrow \mathbf{k}}](\mathbf{k}) = \int_{\mathbb{R}^n} e^{-i\mathbf{k} \cdot \mathbf{x}} f(\mathbf{x}) d\mathbf{x}, \quad (6.4)$$

decomposes the reconstructed image, $f(\mathbf{x})$, into its underlying frequency components. In this work we use two- and three-dimensional discrete FFTs. In this case the extracted image features are the coefficients of the FFT of the μ_a and μ'_s images.

The three-dimensional-FFT ($N_1 \times N_2 \times N_3$ in dimension) results in an $N_1 \times N_2 \times N_3$ matrix of FFT complex coefficients. We truncate the matrix to store only an $n_1 \times n_2 \times n_3$ matrix of coefficients centered at $(N_1 + 1)/2 \times (N_2 + 1)/2 \times (N_3 + 1)/2$, resulting in $n_1 \times n_2 \times n_3$ complex coefficients. Because of the symmetry properties of the FFT and because we are only interested in the absolute value of the coefficients, we reduce the number of distinct coefficients to $(n_1 \times n_2 \times n_3 + 1)/2$. This process allows for the representation of each μ_a and μ'_s image by only $(n_1 \times n_2 \times n_3 + 1)/2$ real-valued coefficients instead of $N_1 \times N_2 \times N_3$ complex FFT coefficients.

In this work, three-dimensional images are parameterized using $n_1 = n_2 = n_3 = 5$, resulting in 63 real-valued coefficients which are labeled from 1 to 63, and ranked based on decreasing distance from the origin. This particular value is chosen because it is optimal in accurately representing the original image and simultaneously maintaining a low coefficient count. The choice of this particular value is based on one assumption: retaining the coefficients of the first few frequencies is sufficient to represent the distribution of optical properties. This assumption is based on the known fact that the light propagation operator is a smoothing operator that does not permit high frequency solutions, a well-known fact in optical tomography. For this reasons we choose to discard high Fourier frequencies. In general, this particular value appears to yield a sufficiently accurate representation of the original image while simultaneously maintaining a low coefficient count.

Table 6.5: Definition of FFT features.

#	Description
13	Absolute error between original image and image captured by the first 5 frequencies of the 2D- or 3D FFT
14-26	(For 2D images only) Absolute value of 2D-FTT coefficients (Fig. 6.6)
14-76	(For 3D images only) Absolute value of 3D-FTT coefficients

Each of the 63 coefficients are treated as independent image features. Similar methodology is used in treating two-dimensional images, which results in 13 unique FFT coefficients for each image. The spatial ordering of three-dimensional FFT coefficients follow the same logic as the ordering scheme for two-dimensional FFT coefficients in Fig. 6.6. The FFT coefficients are labeled according to Table 6.5.

More specifically, projection number two was analyzed with a three-dimensional DFT. The coefficients belonging to the first five frequencies are retained, resulting in a total of 125 coefficients. Since the image $f(\mathbf{x})$ is a real value function, it can be simply obtained that $\hat{f}(\mathbf{k}) = \bar{\hat{f}}(-\mathbf{k})$. Since only the absolute values of the coefficients are stored, these 125 features are reduced to 63, because 62 of them are the complex conjugates of the other 62 features. Thus, 62 features have duplicates and only one is stored. Only features whose corresponding frequency k belonged to the set A_3 are stored (Eq. 6.5).

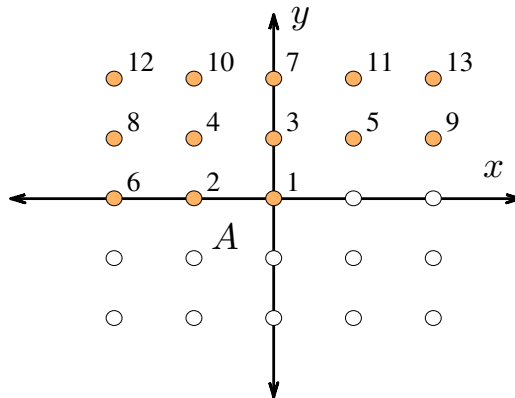


Figure 6.6: FFT coefficients of a two-dimensional image A . Unique coefficients are numbered in increasing order according to distance from the origin, with $z > y > x$ axis preference used as a tie breaker.

$$A_3 \triangleq \{\mathbf{k} : k_1 > 0\} \cup \{\mathbf{k} : k_1 = 0, k_2 > 0\} \cup \{\mathbf{k} : k_1 = 0, k_2 = 0, k_3 \geq 0\} \quad (6.5)$$

Projections 3 to 11 are analyzed with the two-dimensional DFT. Only the coefficients of the first five leading frequencies are stored (total of 25 coefficients). Similar to the three-dimensional DFT, only the 13 features whose corresponding frequency k belong to the set A_2 are stored (Fig. 6.6).

$$A_2 \triangleq \{\mathbf{k} : k_1 > 0\} \cup \{\mathbf{k} : k_1 = 0, k_2 \geq 0\} \quad (6.6)$$

Spectral analysis results in 180 features for each optical property, with 63 features derived from the three-dimensional structured data set and 117 from the two-dimensional projections. Overall, spectral analysis produces 360 features for each finger joint after applying this procedure to the absorption and scattering distributions. We note that this procedure is very flexible and that the number of FFT coefficients that are stored can be increased or reduced as necessary.

6.3.4 Short-hand notation

To succinctly refer to various data sets and extracted features we introduce the following shorthand notation “*Feature #:Projection Name:Optical Parameter.*” For example, the maximum value of the middle sagittal slice in μ_a images is denoted by *F1:GS:a*. Indices “a” and “s” denote μ_a or μ'_s distribution features, respectively. Labeling of FFT features starts with F13 for the first FFT coefficient. For two-dimensional images, the last FFT coefficient is F26, while for three-dimensional images it is F76 (Table 6.5). Feature numbers can be referenced from Tables 6.3, 6.4, and 6.5. Projection names are summarized in Table 6.2.

Combined, features provide information on the distribution of the optical properties inside and around a PIP joint. In total, 55 basic features, 52 GMM parameterization features, and 190 FFT coefficient features are extracted from each finger’s μ_a and μ'_s images (leading to $(55 + 52 + 190) \times 2 = 594$ features).

6.4 Statistical analysis

6.4.1 Kruskal-Wallis ANOVA test and Dunn's test

The utility of each feature for classification is gauged by statistical analysis on the (null) hypothesis that there are no statistically significant differences between the 5 diagnosis groups (A-E) and the control group (H) (Table 6.1). The following three steps are taken to analyze the statistical significance of each feature.

In the first step, through χ^2 goodness of fit analysis we determine that there is only a small likelihood that the extracted features are drawn from a normally distributed population [190]. In the second step, the non-parametric (*distribution-free* or *non-parametric*) Kruskal-Wallis test is used to determine if at least one of the six groups exhibits statistically significant differences from the other groups. Unlike classical ANOVA, the Kruskal-Wallis test is used to determine confidence levels for observed differences between multiple groups through analysis of data rank instead of the actual observations. [190]

The ranking system is as follows: each observation (or feature) is ranked from smallest to largest, without regard to group, with the smallest observation having rank 1. Then, the general Kruskal-Wallis statistic, H , is defined as

$$H = \frac{12}{N(N+1)} \sum_{t=1}^T n_t (\bar{R}_t - \bar{R})^2, \quad (6.7)$$

where \bar{R} is the average rank of all observations, \bar{R}_t is the mean rank observed in group t , and n_t is the size of group t , N is the total number of observations, and T is the total number of groups. The observed differences between the groups are statistically significant if the H -statistic is larger than the corresponding critical value from a χ^2 distribution table with $\nu = T - 1$ degrees of freedom, where $T = 6$, the number of distinct groups [190].

In the third step, group-to-group comparison using Dunn's test is performed to determine which groups, if any, are significantly different from each other. This test is chosen because it allows di-

rect comparison of two groups that do not have the same size [190]. This is of particular importance in our work as group sizes vary significantly (Table 6.1). Dunn’s test is used to compare all possible combinations of two subgroups (i.e. A vs. B, A vs. C, A vs. D, etc.).

Group-to-group comparison using Dunns test is performed to determine which groups are statistically different from each other when observed differences between groups are found statistically significant by the Kruskal-Wallis NOVA test. The Dunns test statistic is defined as,

$$Q = \frac{\bar{R}_a - \bar{R}_b}{\sqrt{\left(\frac{N(N+1)}{12} - \frac{\sum_i^N (\tau_i^3 - \tau_i)}{12(N-1)} \right) \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}}. \quad (6.8)$$

where a and b denote variables from any two groups being compared. Terms involving τ are corrections that account for tied ranks, where τ represents the total number of ties in the i th set of tied ranks.

6.4.2 Effective sample size

Our clinical data consists of 99 fingers from 33 subjects with RA (three fingers per subject) and 120 fingers from 20 subjects without RA (six fingers per subject). In this work we treat each finger as an independent sample. In our calculation of Se and Sp we use the effective sample size (ESS) to account for inter-dependence between DOT images of PIP joints from the same subject (using the intra-class correlation coefficient or ICC) [191, 192]. This procedure reduces the number of independent data samples from 99 (to a minimum of 33) for subjects with RA and from 120 (to a minimum of 20) for subjects without RA, and leads to reduced Se and Sp values.

For each feature we compute the ESS value for the affected (n_A) and healthy (n_H) groups, respectively, and then compute confidence intervals for Se and Sp. The ESS (n) is defined as

$$n = \frac{mk}{1 + \rho(m - 1)}, \quad (6.9)$$

where k is the number of groups or clusters (i.e. subjects), m is the number of samples per group or cluster (i.e. fingers per subject), and ρ is the ICC value, defined as

$$\rho = \frac{s_b^2}{s_b^2 + s_w^2}, \quad (6.10)$$

where s_b^2 is the variance between clusters and s_w^2 is the variance within clusters. The ESS may vary between features depending on the level of correlation between data samples from the same subject as captured by ρ , which in turn may affect the computed Se and Sp values.

The effect of the ICC and ESS on our results is captured by computing the binomial proportion confidence intervals of Se (CI_{Se}^α) and Sp (CI_{Sp}^α) using a Wilson score interval [193]. The confidence intervals are defined as

$$CI_{Se}^\alpha = \frac{Se + \frac{1}{2n_A} Z_{1-\alpha/2}^2 \pm Z_{1-\alpha/2} \sqrt{\frac{Se(1-Se)}{n_A} + \frac{Z_{1-\alpha/2}^2}{4n_A}}}{1 + \frac{1}{n_A} Z_{1-\alpha/2}^2}, \quad (6.11)$$

$$CI_{Sp}^\alpha = \frac{Sp + \frac{1}{2n_H} Z_{1-\alpha/2}^2 \pm Z_{1-\alpha/2} \sqrt{\frac{Sp(1-Sp)}{n_H} + \frac{Z_{1-\alpha/2}^2}{4n_H}}}{1 + \frac{1}{n_H} Z_{1-\alpha/2}^2}, \quad (6.12)$$

where α is the error percentile and $z_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ percentile of a standard normal distribution. For example, to achieve a 95% confidence level, we set $\alpha = 0.05$, so that $1 - \frac{\alpha}{2} = 0.975$ and $z_{1-\frac{\alpha}{2}} = 1.96$. This concept is expanded to a generalized ICC (GICC) when considering Se and Sp for multi-dimensional feature combinations in the second part of this paper. The GICC coefficient is defined as

$$\rho = \frac{\sum_{i,j} |\sigma_{i,j}^b|}{\sum_{i,j} |\sigma_{i,j}^b| + \sum_{i,j} |\sigma_{i,j}^w|}, \quad (6.13)$$

where σ^b and σ^w are the “between” and “within” cluster covariance matrices, respectively [194].

6.4.3 ROC analysis

ROC curve analysis is used to study the classification strength of each feature individually (see Section 5.3.5). For this analysis, we consider only a traditional two-class problem, where one group of data is “positive” and the other group is “negative.” In our case, the group of healthy joints (or group H) is considered “negative” for RA, while the five subgroups that are positive for RA (A, B, C, D, and E) are combined into a single group (group RA).

With ROC curve analysis we determine the classification strength of each feature by computing the threshold value x_{th} that best separates the two groups (i.e. H vs. RA). The best threshold is the feature value x that maximizes the Youden index (Y), which is defined as $Y = Se + Sp - 1$. A feature that perfectly separates the affected from the healthy joints yields $Y = 1.0$, while a feature that completely fails to separate the two classes yields $Y = 0.0$ [185].

6.5 Results

6.5.1 GMM and FFT parameterization

Sample results from GMM parameterization and FFT analysis are presented in Fig. 6.7. The top row corresponds to the middle μ_a cross-sectional slices from the original data. Images in the middle row represent the GMM parameterization of a cross-sectional slice of the original data. Parameterization of the data removes contributions from the boundary, leaving only the major interior structures. In general, the GMM models are good approximations to the original data. Images in the bottom row are reconstructed from only the first five FFT frequencies; they are

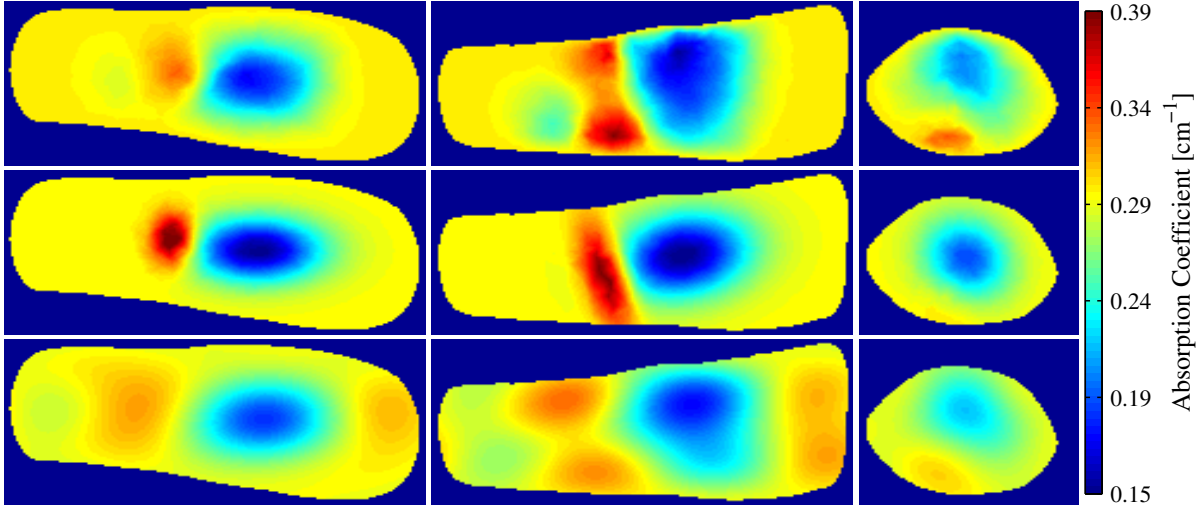


Figure 6.7: (Top row) Sagittal, coronal, and transverse cross sections of a healthy PIP joint. (Middle row) GMM parameterization approximation. (Bottom row) Reconstruction of the original image using only the extracted FFT coefficients of the first five frequencies.

representative of the level of detail captured by the FFT coefficients we extract, capturing the general distribution of μ_a . Preserving the first five frequencies minimizes the contribution from pixels near the boundary; this is important because values near the boundary are more prone to numerical error and noise. Similar results are found for the μ'_s data.

6.5.2 Kruskal-Wallis ANOVA test

Results from Kruskal-Wallis analysis of features from images of PIP joints from groups A-H are summarized in Fig 6.8. We plot the H -statistic as a function of data set and feature number (see Sections 6.2- 6.3; Tables 6.2, 6.3, 6.4, and 6.5). There are six distinct groups ($k = 6$), thus, $H > 11.07$, 15.09, and 20.52 are necessary to establish statistical significance in observed difference at the 0.05, 0.01, and 0.001 confidence level; $H \geq 11.07$ in 249 features (131 from μ_a and 118 from μ'_s) and $H \geq 20.52$ in 129 features (55 from μ_a and 74 from μ'_s).

Basic features (two-dimensional and three-dimensional) and features from the two-dimensional-FFT of the VS and GS slice of μ_a data yield the most features with $H > 20.52$. In the case of μ'_s images, basic features (two-dimensional and three-dimensional) and two-dimensional-FFT coefficients of projections SC , VS , VT , VC , and GS result in many feature with $H > 20.52$.

The results show statistically significant differences between the spectral features of PIP joints of subjects with RA and without RA.

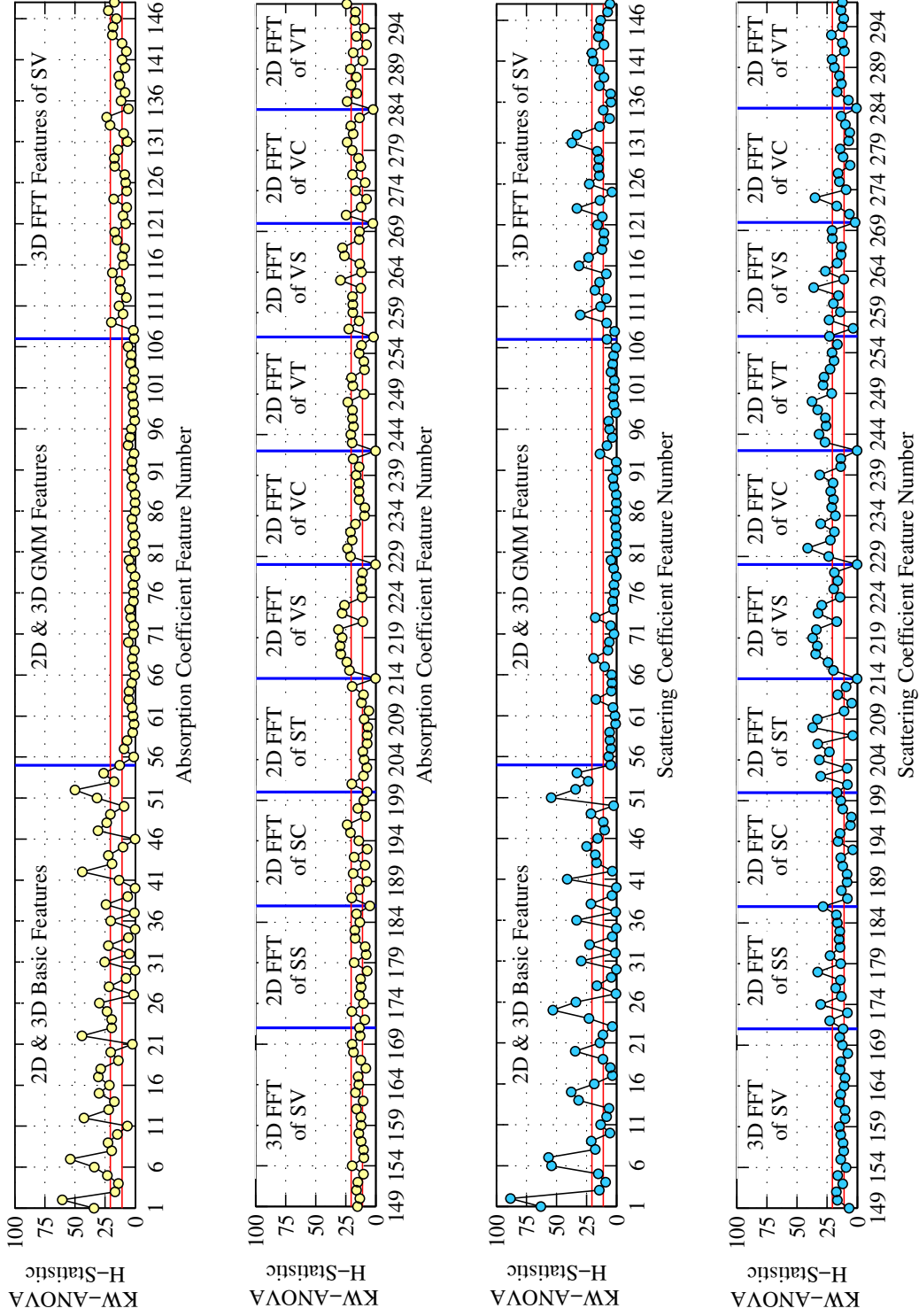


Figure 6.8: H -statistic value for μ_a (top two rows) and μ_s (bottom two rows) features. The threshold values for H at the 0.05 and 0.001 confidence levels are 11.07 and 20.52, respectively, and shown as horizontal red lines.

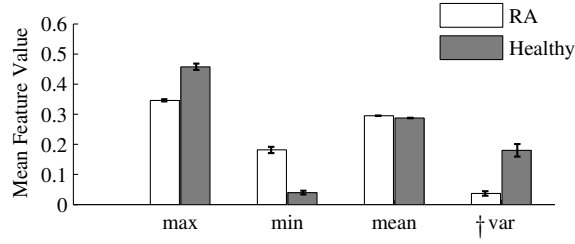
6.5.3 Dunn's test

Table 6.6 shows results from Dunn's test applied to the first five coefficients of the three-dimensional-FFT of the structured data, corresponding to the low frequency components of the μ_a and μ'_s distributions. The critical Q values to establish statistically significant differences between two subgroups at the 0.05 and 0.01 significance levels are $Q = 2.936$ and $Q = 3.403$, respectively. Instances where $Q > 2.936$ are highlighted in bold.

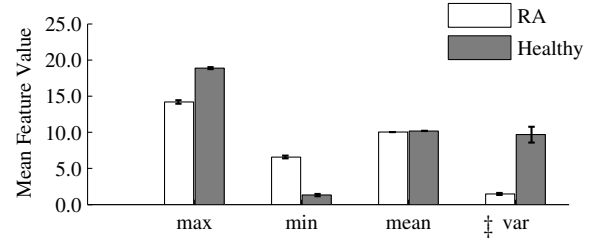
One can see that Q is generally greater than 2.936 when features of group H are compared to each of the other affected groups (A, B, C, D, E). This means that these features may be useful in distinguishing between healthy volunteers and subjects with RA. On the other hand, Q is generally smaller than 2.936 when affected groups are compared to each other; indicating that it may be difficult to distinguish between the different subgroups of affected subjects. However, there are some features with $Q > 2.936$, such as *F14:SV:a* and *F14:SV:s* (groups B vs. E), suggesting that even these subgroups may be distinguishable. Similar results are obtained for all other features.

Figure 6.9 shows the mean values and standard errors of the maximum ($n_H = 59.9$, $n_A = 52.1$), minimum ($n_H = 45.1$, $n_A = 53.6$), mean ($n_H = 55.1$, $n_A = 52.0$), and variance ($n_H = 62.1$, $n_A = 52.0$) of μ_a and μ'_s images for healthy volunteers and subjects with RA. The standard error was computed using the effective sample size of the affected and healthy groups (Eq. 6.9), denoted by n_H and n_A for each specific feature, respectively. One can see that, on average, the healthy subjects show a higher maximum μ_a value and a higher variance. On the other hand, the minimum and mean μ_a values are lower in healthy subject as compared to subjects with RA.

Similar to observations from μ_a images, subjects with RA have a lower maximum μ'_s value but a higher minimum μ'_s value compared to healthy subjects. However, in contrast to results from μ_a data, subjects with RA have a marginally lower mean μ'_s value. Similar to results from μ_a images, subjects with RA have a significantly lower variance in μ'_s images when compared to healthy subjects.



(a) Absorption Coefficient Features



(b) Scattering Coefficient Features

Figure 6.9: Mean value and standard error of the maximum, minimum, mean, and variance of μ_a and μ'_s images. A two-sample *student-t* test shows differences between features from subjects with RA and without RA are statistically significant at the $\alpha = 0.01$ level. The variance is scaled to display on the same axis († scaled by 100; ‡ scaled by 10).

Table 6.6: Dunn's test for sample features, showing statistically significant differences between subgroups.

Parameter	Feature #	H vs. A	H vs. B	H vs. C	H vs. D	H vs. E	A vs. B	A vs. C	A vs. D	A vs. E	B vs. C	B vs. D	B vs. E	C vs. D	C vs. E	D vs. E
Absorption Coefficient	<i>F13:SV:a</i>	5.44	6.39	4.99	7.32	5.94	0.72	0.37	0.61	0.15	0.27	0.18	0.92	0.14	0.52	0.83
	<i>F14:SV:a</i>	4.21	0.89	3.99	3.78	5.49	2.52	0.39	0.85	0.53	2.64	1.90	3.22	1.16	0.06	1.51
	<i>F15:SV:a</i>	2.97	4.80	3.47	5.62	3.25	1.39	0.81	1.47	0.07	0.44	0.06	1.56	0.42	0.91	1.67
	<i>F16:SV:a</i>	2.90	5.58	3.49	5.53	3.86	2.03	0.87	1.46	0.42	0.95	0.76	1.75	0.35	0.55	1.12
	<i>F17:SV:a</i>	2.02	4.86	2.17	4.94	3.48	2.15	0.39	1.78	0.86	1.54	0.58	1.44	1.14	0.35	0.98
Scattering Coefficient	<i>F13:SV:s</i>	6.29	6.39	5.28	7.90	6.91	0.07	0.03	0.30	0.14	0.04	0.22	0.22	0.24	0.15	0.48
	<i>F14:SV:s</i>	1.73	2.18	1.63	0.23	3.25	2.97	0.15	1.28	0.93	2.81	1.98	4.10	1.28	0.66	2.41
	<i>F15:SV:s</i>	5.32	4.48	5.36	7.69	6.32	0.64	0.75	0.97	0.22	1.32	1.66	0.90	0.04	0.59	0.80
	<i>F16:SV:s</i>	5.06	3.18	2.62	5.56	4.60	1.43	1.30	0.31	0.80	0.03	1.25	0.72	1.12	0.66	0.55
	<i>F17:SV:s</i>	4.33	4.05	2.61	5.94	3.67	0.21	0.82	0.56	0.88	0.63	0.79	0.65	1.37	0.09	1.59

6.5.4 ROC analysis

Examples of ROC curves are given in Fig.6.10, while in Fig. 6.11 we show the results from ROC curve analysis of μ_a and μ'_s features. We plot the Youden index (Y) as a function of the data set and the feature number (see Sections 6.2- 6.3; Tables 6.2, 6.3, 6.4, and 6.5). $Y \geq 0.7$ for 107 features, where 65% are from μ'_s and 35% from μ_a images. $Y \geq 0.80$ for three μ'_s features; variance of unstructured data (0.82), mean of variance between transverse slices (0.81), and the first FFT coefficient (i.e. DC component) of the variance between coronal slices (0.80). The largest Y from μ_a images is obtained with the ratio of maximum to minimum of the summation of all transverse slices (0.77). In general, the best single feature classification results are from μ'_s features.

Of the 107 features that achieve $Y \geq 0.70$, approximately 50% were basic statistical features, 45% from spatial Fourier analysis, and 5% from GMM parameters. Over 47% of the features resulting in $Y \geq 0.70$ are derived from the variance across two-dimensional sagittal, transverse, and coronal planes (51 of 107 features).

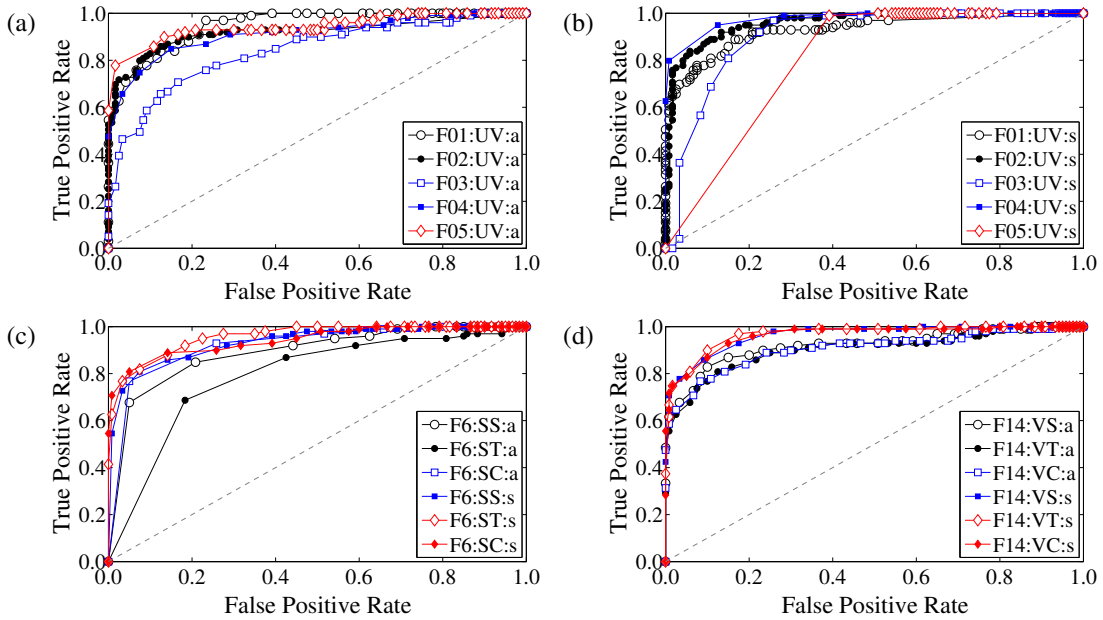


Figure 6.10: (a-b) ROC curves for basic features from unstructured data (F01:UV:a and F01:UV:s, F02:UV:a and F02:UV:s, etc.), (c) absolute error between original data and GMM model features, and (d) DC component of two-dimensional-FFT features. FFT features perform better, in terms of area under the curve, than basic features and GMM coefficients.

Basic features from μ'_s images are better classifiers than features from μ_a images (Fig.6.10(a-b)). The variance of the unstructured data ($F4:UV:s$) is the best single feature classifier ($Y = 0.82$) (Fig.6.10(b)). The absolute error between original images and GMM approximations performed strongly as one-dimensional classifiers with up to $Y = 0.75$ (Fig.6.10(c)). The coefficients of the lowest order term of the two-dimensional-FFT of variance images (VS , VT , VC) are very strong one-dimensional classifiers with up to $Y = 0.80$ (Fig.6.10(d)).

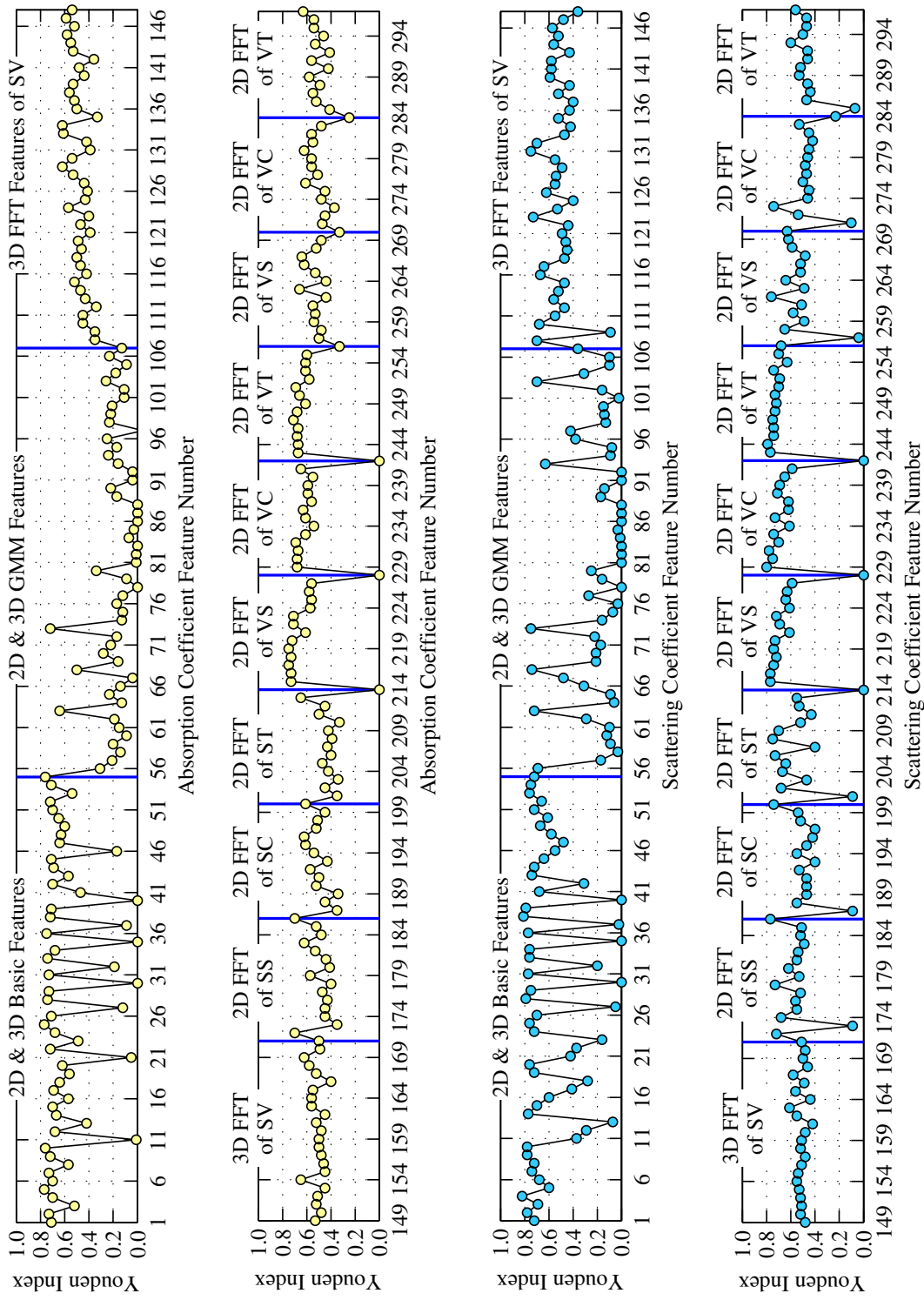


Figure 6.11: Single feature classification results reported as the Youden index ($Se + Sp - 1$) from ROC analysis for μ_a (top two rows) and μ'_s (bottom two rows) features.

6.6 Discussion

In this chapter we presented a method for extracting features from DOT images that goes beyond simple statistical values. The process consists of three steps: (1) data pre-processing, (2) feature extraction, and (3) statistical and ROC curve analysis of individual features.

The framework is tested on 219 DOT images of PIP joints II-IV gathered through a clinical trial (33 subjects with RA and 20 healthy control subjects). Clinical diagnosis of RA according to the ACR criteria is the gold standard. Ultrasound and MRI scans of the clinically dominant hand were performed. A rheumatologist and a radiologist, in a blinded review, analyzed the images and classified each subject based on detectable symptoms of RA (groups B to E). Subjects without signs of RA-induced joint deformities were classified as healthy (group H) or as affected with RA (group A) based on non-imaging based evidence. A total of 297 features are extracted from each μ_a and μ'_s reconstruction of all imaged joints. Statistical analysis of the extracted features is used to find features that reveal statistically significant differences between diagnosis groups.

Three important findings are discovered. First, through application of the non-parametric Kruskal-Wallis ANOVA test we establish the existence of image features that show statistically significant differences between subjects with RA and without RA.

Second, we use Dunn's test ($p < 0.05$) to discover that features derived from group A (subjects with RA but without abnormal findings in MRI and ultrasound scans) are statistically different from features derived from healthy subjects (group H). At the same time, the differences in optical properties between group A and groups B to E were generally not significant. This is an important finding because it shows that DOT imaging of PIP joints has the potential to detect the presence of RA even when ultrasound and MRI scans cannot detect effusion, synovitis, or erosion in the joint cavity.

Third, we discover that the μ'_s distribution yields stronger one-dimensional classifiers. ROC analysis shows three μ'_s features for which $Y \geq 0.8$. This is a significant improvement over previous work where $Y \leq 0.59$ was obtained with ROC analysis [9]. This represents the first time that μ'_s images are exploited to yield strong differences between PIP joints of subjects with and

without RA. Additionally, we establish that the variance between sagittal, transverse, and coronal slices yields a large number of strong single feature classifiers ($Y \geq 0.70$).

These three findings demonstrate that changes in optical properties induced by RA are detectable using DOT. The statistically significant differences between image features from affected and healthy subjects shows that it might be possible to accurately diagnose RA using FD-DOT. Furthermore, the general lack of statistically significant differences between features from group A and groups B to E is evidence that DOT is sensitive to changes in optical properties of the synovium that MRI and ultrasound cannot resolve.

The following chapter, Chapter 7, focuses on multi-dimensional classification using the features that result in the largest Youden indices from ROC analysis as presented in this chapter. Employing five different classification algorithms, we show that using multiple rather than single image features leads to higher sensitivities and specificities.

Image Classification

The content of this chapter advances the research in Chapter 6 by employing data mining techniques to achieve high sensitivities and specificities. Our efforts in prior work and those in Chapters 5 and 6 employ only basic feature extraction schemes and basic image classification techniques. Those results show that there is a significant level of natural contrast in the optical properties of PIP joints, likely arising from the onset of RA, indicating that DOT is a promising technique for diagnosing RA. Results from this chapter further demonstrate the utility of CAD in enhancing our ability to diagnose RA from DOT images.

In Chapter 6 we present a framework for processing DOT images and extracting heuristic features. The classification strength of each feature is evaluated with Kruskal-Wallis ANOVA, Dunn's test, and ROC curve analysis. Three important observations are made. First, we observe that features of subjects with RA differ from features of healthy subjects ($p < 0.05$). This implies that physiological differences between subjects with RA and healthy subjects can be captured by DOT images. Our second major finding pertains to subjects with RA that do not exhibit effusions, synovitis, or erosion on MRI and ultrasound scans. The DOT images of these subjects are statistically the same as the images of subjects with RA that do exhibit effusions, synovitis, or erosion; suggesting that DOT can potentially detect the onset of RA in these joints before MRI and US show macroscopic anatomical changes. Our third major finding shows that single features from μ'_s images allow more accurate one-dimensional classification (i.e. using ROC curve analysis) of each

joint as affected or not affected by RA compared to features from μ_a images.

In this chapter we present a general framework for classifying DOT images of PIP joints as affected or not affected by RA using machine learning techniques that allow the use of multiple features in the analysis; thus going beyond the ROC curve analysis of Chapter 6, which we limit to the evaluation of one feature at a time. Classification of each PIP joint is performed with the best 30 features from Chapter 6 and five distinct classification algorithms: (1) linear discriminant analysis (LDA), (2) quadratic discriminant analysis (QDA), (3) k -nearest neighbors (KNN), (4) self-organizing maps (SOM), and (5) support vector machines (SVM). We report the performance of each algorithm in terms of sensitivity (Se) and specificity (Sp).

The work in this paper goes beyond previously published material [9, 140, 173] in several ways. (1) First, unlike in previous studies, we combine μ_a and μ'_s features in the classification analysis. (2) Second, we are substantially increasing the number of features considered from 4 to 594. In addition to basic features, such as smallest or largest μ_a and μ'_s values, we now consider more advanced features, such as Fourier coefficients of two-dimensional and three-dimensional spatial distribution of optical properties. (3) Third, we compare the performance of five different classification algorithms to determine which scheme is most suitable for DOT imaging data. (4) Fourth, we employ a feature-selection algorithm to determine the subset of image features that achieves highest Se and Sp. This step is essential given the large number of permutations possible when 594 features are considered together with five different classification algorithms. Here we employ an evolution strategy that finds 5 to 10 optimal features out of 594. (5) Finally, we use intra-cluster correlation coefficients (ICC) to compute the effective sample size (ESS) of each data group, which we then use to adjust the Se and Sp values; this helps to account for bias that may arise as a result of treating each imaged finger as an independent sample. This step allows us to compute confidence intervals for Se and Sp, a necessary treatment as our data consists of multiple fingers per subject and these images may not be statistically independent.

In the remainder of this chapter we address the general format of multi-dimensional classification by presenting details on the five classification methods of interest. We review the mathematical

and theoretical foundations of the feature-selection algorithm and present results from classification of DOT images of PIP joints as affected or not affected with RA. All classification results are validated through extensive cross-validation. The chapter concludes with a discussion on the potential impact of CAD in the diagnosis of RA with DOT.

7.1 Data

As was previously mentioned, the analysis in this chapter is focused on a subset of the features obtained in Chapter 6. Reducing the number of features from the original 594 features to a smaller subset is motivated by various factors. (1) Features that poorly differentiate between the two diagnostic groups in ROC analysis are unlikely to offer substantial contributions in multi-dimensional analysis and should be discarded. (2) To improve the generalizability of classification results it is generally desirable for the ratio between the number of features (l) and data samples (N) to be small (generally $l/N = 0.1$ to $l/N = 0.2$ are acceptable) [189]. (3) The complexity of classification algorithms can increase with l , sometimes exponentially. Ultimately, the final number of features should strike a balance between these motivating factors and the desire to include as many important features as possible.

In this work we select 30 features from the original 594 features, as it is a good compromise, resulting in $l/N = 0.13$. These 30 features are selected based on Youden indices ($Y = \text{Se} + \text{Sp} - 1$) from ROC curve analysis. The features with the 30 largest Y values are selected, of these, 4 are from μ_a and 26 are from μ'_s images. Throughout this paper we refer to features by the numbering in Table 7.1.

The original clinical data is divided into five groups (labels A, B, C, D, or E) of subjects with RA (segmented based on symptoms detected on MRI and US scans) and one group without RA (label H) (refer to Table 5.1). We establish in Chapter 6 that the subgroups of affected subjects are not statistically different from each other based on the features we currently consider. However, each of the affected subgroups is statistically different from the cohort of healthy subjects. As a

Table 7.1: List of features with top Youden indices from ROC curve analysis (presented in Chapter 6).

#	Descriptive Notation	Analysis Type	Optical Variable	#	Descriptive Notation	Analysis Type	Optical Variable	#	Descriptive Notation	Analysis Type	Optical Variable
1	F05:UV:a	Basic	Absorption	11	F05:ST:s	Basic	Scattering	21	F013:SC:s	2D-FFT	Scattering
2	F05:SV:a	Basic	Absorption	12	F03:VS:s	Basic	Scattering	22	F014:VS:s	2D-FFT	Scattering
3	F05:ST:a	Basic	Absorption	13	F01:VC:s	Basic	Scattering	23	F015:VS:s	2D-FFT	Scattering
4	F05:GT:a	Basic	Absorption	14	F03:VC:s	Basic	Scattering	24	F014:VC:s	2D-FFT	Scattering
5	F02:UV:s	Basic	Scattering	15	F04:VC:s	Basic	Scattering	25	F015:VC:s	2D-FFT	Scattering
6	F04:UV:s	Basic	Scattering	16	F01:VT:s	Basic	Scattering	26	F016:VC:s	2D-FFT	Scattering
7	F04:SV:s	Basic	Scattering	17	F03:VT:s	Basic	Scattering	27	F014:VT:s	2D-FFT	Scattering
8	F05:SV:s	Basic	Scattering	18	F04:VT:s	Basic	Scattering	28	F015:VT:s	2D-FFT	Scattering
9	F04:SS:s	Basic	Scattering	19	F03:GT:s	Basic	Scattering	29	F018:VT:s	2D-FFT	Scattering
10	F05:SC:s	Basic	Scattering	20	F04:GT:s	Basic	Scattering	30	F019:GS:s	2D-FFT	Scattering

result, in the following analysis we treat all subjects with RA as one group (affected with RA) and attempt to accurately classify an unseen data point as affected or not affected with RA.

The nomenclature used in Table 7.1 is established in Chapter 6 and follows the pattern “*Feature # : Projection Name : Optical Parameter.*” Shorthand notation is necessary because of the large number of features we consider in Chapter 6. For example, feature 19 is denoted by F03:GT:s, which translates to the mean (F03 or feature number 3) of the geometrically dominant transverse (GT) slice of the reduced scattering (s) reconstruction. Indices “a” and “s” denote μ_a or μ'_s derived features, respectively. Feature numbers F01 to F05 are basic statistical features, F06 to F12 are Gaussian mixture model features, and labeling of FFT features starts from F13 for the first FFT coefficient. For two-dimensional images, the last FFT coefficient is F26, while for three-dimensional images it is F76.

As an example and to visualize the reduction of subgroups A-E into a single group, consider features 5 and 19 from Table 7.1. The allocation of data to subgroups A-E and H is presented in Fig. 7.1(a), where all six groups are visualized. The same data is presented in Fig. 7.1(b); however, in this plot, a single group (label RA) replaces subgroups A-E, resulting in only two groups of data. For the purpose of consistency, we use features 5 and 19 throughout this text when it is necessary to show two-dimensional plots.

In this work we study the classification performance of various multi-dimensional combinations of these 30 features, starting with two-dimensional combinations. In Sections 7.2.1-7.2.4 we

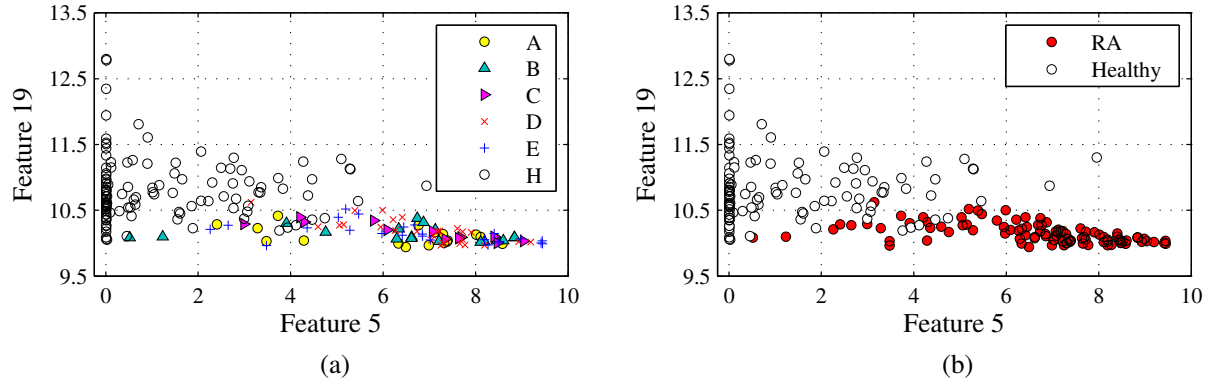


Figure 7.1: Distribution of sample features. In (a) the six distinct diagnosis groups (A, B, C, D, E, H) are identified, while in (b) the five cohorts diagnosed with RA (A, B, C, D, E) are grouped into one “RA” group.

briefly review the five classification algorithms. The cross-validation methodology is presented in Section 7.2.5. The feature selection algorithm, which we use to find the optimal feature combination, is presented in Section 7.2.6.

7.2 Methods

7.2.1 The KNN method

The KNN algorithm is among the most basic classification algorithms because it does not require much, if any, prior knowledge of the distribution of the training or testing data. Each unseen feature vector, \mathbf{x} , is classified according to the density of affected or healthy data points within a spatial sphere of radius r (covering k neighboring data points and two distinct data classes M) [189, 195]. The rules governing which label is assigned to each testing vector \mathbf{x} are:

1. From the training data, identify the k nearest neighbors to each vector \mathbf{x} using the Euclidean distance measure.
2. Count the number of training data vectors belonging to each class (Healthy or RA).
3. Assign test vector \mathbf{x} to the class with the maximum number of k_i samples (Healthy or RA).

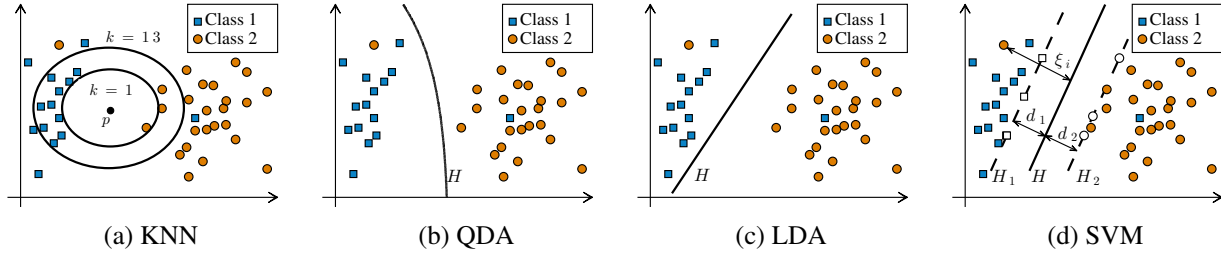


Figure 7.2: (a) KNN, (b) QDA, (c) LDA, and (d) SVM boundaries for a non-linearly separable two-class problem. (a) Examples of $k = 1$ and $k = 13$ for KNN are shown for a new data point p , denoted by \blacktriangle . (d) Support vector data points from classes 1 and 2 are denoted by white square or circle markers, respectively.

The choice of k affects classification: generally, larger values reduce the effect of noise, but make boundaries between classes less distinct (Fig.7.2(a)). The simplest version of the algorithm is when $k = 1$, known as the nearest neighbor (NN) rule. In other words, a feature vector \mathbf{x} is assigned to the class of its nearest neighbor.

7.2.2 The LDA and QDA methods

Classification with discriminant analysis (DA) is a popular parametric method based on Bayesian statistics, primarily used when the training and testing data is believed to be normally distributed. Even in cases where the data is not normally distributed, DA is generally an effective classifier as it gives rise to linear and quadratic hyperplanes that are reasonably accurate at separating the two classes.

References for the theoretical development and implementation of DA can be found in [189, 196]. In general, for classification with discriminant analysis (DA), the posterior probability $p(\omega_i|\mathbf{x})$ of feature vector \mathbf{x} originating from class ω_i is defined by *Bayes theorem* [187, 189, 195],

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i) P(\omega_i)}{P(\mathbf{x})}, \quad (7.1)$$

where $P(\omega_i)$ and $P(\mathbf{x})$ are prior probabilities for class ω_i and feature vector \mathbf{x} , respectively [189]. Classification of \mathbf{x} is done using the maximum a posteriori estimate, $\max_{\omega_i} \hat{p}(\omega_i|\mathbf{x})$, and setting the prior probability for each feature vector equal, $\hat{p}(\omega_i|\mathbf{x}) \propto \hat{p}(\mathbf{x}|\omega_i)P(\omega_i)$.

The prior probabilities for each class is defined to be equal, $P(\omega_i) = P(\omega_j) \forall i, j$, so classification depends only on the likelihood function estimate, $\hat{p}(\mathbf{x}|\omega_i)$. The likelihood functions are assumed to follow the general multivariate normal distribution,

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{l/2}|\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right], \quad i = 1, \dots, M, \quad (7.2)$$

where l is the dimensionality of \mathbf{x} , M the number of classes (here $M = 2$), μ_i the mean value, and Σ_i the covariance matrix of class ω_i . For classification purposes, estimates for μ_i and Σ_i are computed from the training data, using the maximum likelihood estimation. Following Bayes, classification is performed using the discriminant function

$$g_i(x) = \ln(p(x|\omega_i)P(\omega_i)), \quad (7.3)$$

where $P(\omega_i)$ is the *a priori* probability of class ω_i . Here, the assumption that $P(\omega_i) = P(\omega_j) \forall i, j$ is used. The decision surfaces resulting from the discriminant functions are $g_i(x) - g_j(x) = 0$.

Two distinct classification methodologies arise from this theory. In the first and more general case, the covariance matrix for each class is estimated independently, resulting in hyperquadratic decision curves (Fig. 7.2(b)). This method is generally referred to as quadratic discriminate analysis (QDA). In the second case, the covariance matrices of the two groups are assumed to be identical ($\Sigma_i = \Sigma_j \forall i, j$). The resulting decision curves are hyperplanes and the method is called linear discriminate analysis (LDA) (Fig. 7.2(c)). In both cases, the individual features are assumed to be statistically independent, resulting in diagonal covariance matrix estimates [189].

More specifically, the DA method creates a “decision boundary” that allows for the classification of a given data point (image) into one of the two classes: healthy or affected. The classification is made based on the location of a data point relative to the decision boundary. The calculation of the decision plane in DA relies on the distance between a given data point and the class centroids. For example, in the two-dimensional case, where we consider a combination of two image

parameters, QDA produces a quadratic (Fig. 7.2b) while LDA produces a line as the decision plane (Fig. 7.2c). Increasing the number of dimensions (i.e. number of image parameters considered) increases the dimensionality of the decision hyperplanes.

7.2.3 The SOM method

The SOM algorithm is a type of constrained clustering algorithm, where an initial set of randomly distributed and discrete points (“neurons”) are allowed to “self-organize” into a smooth manifold. The self-organizing process is achieved through training, a type of competitive learning process, and is typically referred to as vector quantization. After “clustering” is complete, each neuron is assigned a class label (Healthy or RA) based on the number of training vectors from each training class (in this way, similar to the KNN algorithm). Finally, the testing data is input and each feature vector \mathbf{x} in the testing set is assigned to its topologically corresponding neuron. The test vector \mathbf{x} therefore inherits the class label of its assigned neuron.

Our team of researchers previously presented the theoretical developments necessary for the application of SOMs to CAD of DOT images [173]. In this work we use SOMs to perform image classification in multi-dimensional feature space, varying the number of neurons (n) and learning rate (l) used for pattern learning. The total neurons are varied between 9, 16, and 25. The learning rate is varied from 0.01, 0.1, and 1.0.

7.2.4 The SVM method

SVM has become very popular in the field of machine learning over the past two decades and is increasingly popular because of its utility in classification and pattern recognition. It has been shown to have very good generalization properties, and unlike statistical classification done using discriminate analysis (DA), SVM makes no assumption on the underlying distribution of the sample data.

We use SVM for the general two-class problem, where the classes are not necessarily linearly separable (Fig. 7.2d and Fig. 7.3b) [189, 195, 197]. We review the well-established SVM theory

for completeness. We begin with the theory for linearly separable problems. We then present the theory for non-separable problems.

Linearly separable problem

SVM is a quadratic optimization problem, where the line (or hyperplane in arbitrary dimensions) that best separates the two classes (+ or -) is chosen so that the margin between the support vectors from each class is maximized (Fig. 7.3a). The optimal separating line (hyperplane in n -dimensional space), H , is denoted $H \rightarrow \omega^T x + b = 0$. Then, the following terms can be defined without loss of generality.

$$H \rightarrow \omega^T x + b = 0 \quad (7.4)$$

$$H^+ \rightarrow \omega^T x + b = +1 \quad (7.5)$$

$$H^- \rightarrow \omega^T x + b = -1 \quad (7.6)$$

The distance from the origin to the separating hyperplane is given by,

$$q = \min_x \|\mathbf{x} - 0\|_2^1 \quad \text{subject to} \quad \omega^T x + b = 0. \quad (7.7)$$

The value of q can be solved using a Lagrange multiplier, by posing the problem as the following optimization program,

$$\min_x \left(\|\mathbf{x} - 0\|_2^1 + \lambda (\omega^T x + b) \right). \quad (7.8)$$

The value of q is minimized when $x = -\lambda\omega$. Then, using the given condition, $\omega^T \mathbf{x} + \mathbf{b} = 0$, it is clear that the value of the Lagrange constant is

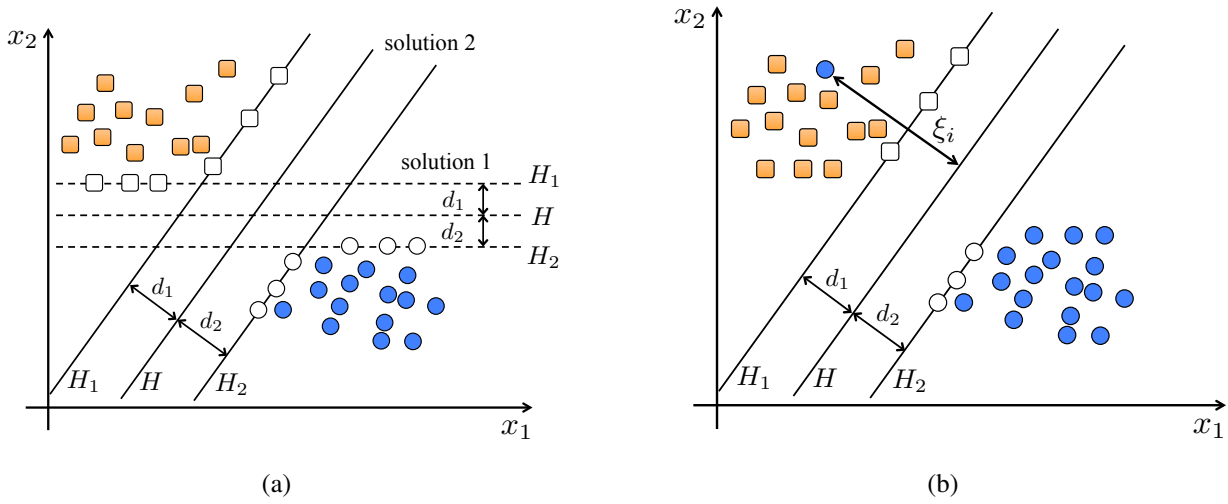


Figure 7.3: (a) Example of linearly separable two-class problem that can be separated by infinite different lines, but there is only one optimal separating line. The “margin” corresponding to each acceptable solution (separating line) is denoted with dashed lines. In this example, solution 2 is preferred over solution 1 because of the size of the margin. (b) Example of non-linearly separable two-class problem. The “slack” variable ξ is highlighted.

$$\lambda = -\frac{b}{\omega^T \omega}. \quad (7.9)$$

This result suggests that the distance from the origin to the hyperplane is

$$q = \frac{|b|}{\|\omega\|}. \quad (7.10)$$

Then, the constraints on the SVM for the linearly separable two-class problem are,

$$\omega^T x_i + b \geq +1 \quad \forall y_i = +1, \quad (7.11)$$

$$\omega^T x_i + b \leq -1 \quad \forall y_i = -1, \quad (7.12)$$

or equivalently,

$$y_i (\omega^T x_i + b) - 1 \geq 0. \quad (7.13)$$

This allows the useful definition of the distance from the origin to the hyperplane, from the origin to the positive margin boundary, and from the origin to the negative margin, respectively,

$$H \rightarrow q = \frac{|b|}{\|\omega\|}, \quad (7.14)$$

$$H^+ \rightarrow q_+ = \frac{|b - 1|}{\|\omega\|}, \quad (7.15)$$

$$H^- \rightarrow q_- = \frac{|b + 1|}{\|\omega\|}. \quad (7.16)$$

Then, the margin of the SVM can be defined as,

$$m = d_+ + d_- = \frac{2}{\|\omega\|}. \quad (7.17)$$

This is an important property of SVM. It states that *maximizing* the margin m is equivalent to *minimizing* $\|\omega\|$. As a direct result, the primal SVM quadratic optimization problem can be stated as

$$L_p : \arg \min_{\omega} \frac{1}{2} \|\omega\|^2 \quad \text{subject to} \quad y_i (\omega^T x_i + b) - 1 \geq 0, \quad (7.18)$$

where ω denotes the best separating hyperplane (i.e. $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$), and y_i is the class label of the i^{th} data feature vector. This can once again be re-written as an optimization problem using Lagrange multipliers,

$$\min_{\omega, b} \max_{\alpha \geq 0} \left(\frac{1}{2} \|w\|^2 - \sum_i \alpha_i (y_i (w^T x_i + b) - 1) \right). \quad (7.19)$$

This can be re-written in its dual form by solving for $\frac{\partial L_p}{\partial \omega}$ and $\frac{\partial L_p}{\partial b}$ and plugging it into the L_p . The resulting quadratic program is the dual SVM (no longer a mixture of minimization and maximization), and is written as

$$L_D : \max_{\alpha \geq 0} \left(\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j \right) \quad \text{subject to} \quad \sum_i \alpha_i y_i = 0, \quad \alpha_i \geq 0. \quad (7.20)$$

Classification is done using the following scheme

$$f(x) = \text{sign}(\omega^T x + b) \quad (7.21)$$

$$= \text{sign} \left(\sum_i \alpha_i y_i x^T x_i + b \right), \quad (7.22)$$

where the value of b is found from the *non-zero* α 's.

$$\tilde{b} = y_i - \omega^T x_i \quad \forall \alpha_i > 0 \quad (7.23)$$

$$b = \text{average}(\tilde{b}) \quad (7.24)$$

Thus, any data point (or feature vector) x can be classified using the above scheme, where $f(x) = \pm 1$ states whether data point x is classified into class 1 or -1 .

Non-separable problem

The SVM formulation must be altered for the case of non-separable two-class problems (Fig. 7.3b). This is achieved by introducing a “slack” variable ξ , which allows for points to fall on the “wrong” side of the separating plane (i.e. misclassified points). The new constraints are,

$$\omega^T x_i + b \geq +1 - \xi_i \quad \forall y_i = +1, \quad \xi_i \geq 0, \quad (7.25)$$

$$\omega^T x_i + b \leq -1 + \xi_i \quad \forall y_i = -1, \quad \xi_i \geq 0, \quad (7.26)$$

or, simply

$$y_i (\omega^T x_i + b) - 1 + \xi_i \geq 0, \quad \xi_i \geq 0. \quad (7.27)$$

Slack values of $\xi_i = 0$ mean that data x_i is correctly classified (it is on the correct side of the decision boundary). The primal SVM problem for the non-separable two-class problem is

$$L_p : \arg \min_{\omega} \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i \quad \text{subject to} \quad y_i (\omega^T x_i + b) - 1 + \xi_i \geq 0. \quad (7.28)$$

The variable C is a cost variable that penalizes the objective function for points that are misclassified. The case where $C \rightarrow 0$ corresponds to the case where misclassified points do not adversely affect the objective function. The case where $C \rightarrow \infty$ corresponds to the case where misclassified points severely adversely affect the objective function. The non-separable SVM corresponds to the case where $C \rightarrow \infty$. The primal SVM can be written using Lagrange multipliers,

$$\min_{\omega, b, \xi} \max_{\alpha \geq 0, \beta \geq 0} \left(\frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i - \sum_i \alpha_i (y_i (\omega^T x_i + b) - 1 + \xi_i) - \sum_i \beta_i \xi_i \right). \quad (7.29)$$

By taking the partial derivatives of the objective function with respect to ω , b , and β_i we obtain,

$$\frac{\partial L_p}{\partial \omega} = 0 = \omega - \sum_i \alpha_i y_i x_i, \quad (7.30)$$

$$\frac{\partial L_p}{\partial b} = 0 = - \sum_i \alpha_i y_i, \quad (7.31)$$

$$\frac{\partial L_p}{\partial \xi_i} = 0 = C - \alpha_i - \beta_i. \quad (7.32)$$

This implies that $\alpha_i = C - \beta_i$. Given that $\alpha_i \geq 0$ and $\beta_i \geq 0$, these results suggest the following constraints on the Lagrange multipliers

$$0 \leq \alpha \leq C, \quad (7.33)$$

$$\beta_i \geq 0. \quad (7.34)$$

The dual SVM for the non-separable two-class case can then be written as:

$$L_D : \max_{\alpha \geq 0} \left(\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j \right) \quad \text{subject to} \quad \sum_i \alpha_i y_i = 0, \quad \alpha_i \in [0, C]. \quad (7.35)$$

This quadratic program is solved for α_i . Subsequently, the optimal separating hyperplane can be recovered from one of the constraints, $\omega = \sum_i \alpha_i y_i x_i$, and classification is done using the scheme $f(x) = \text{sign}(\omega^T x + b)$ or $f(x) = \text{sign}\left(\sum_i \alpha_i y_i x^T x_i + b\right)$.

The value of b is found from only those data vectors that are on the margin (i.e. the support vectors, Fig. 7.2(d)). The support vectors correspond to vectors with non-zero and not- C α 's and by assuming $\xi_i = 0$. Then, $b = \tilde{b}$, where b is computed from $y_i (\omega^T x_i - \tilde{b}_i) - 1 + \xi_i = 0$.

Thus, any data point (or feature vector) x can be classified using the above scheme, where

$f(x) = \pm 1$ states whether data point x is classified into class $+1$ or -1 .

Non-linear SVM

In general, the hyperplane that separates the two classes does not need to be linear. The dual SVM for the non-separable two-class case can be written as

$$L_D : \max \left(\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j) \right) \quad \text{subject to} \quad \sum_i \alpha_i y_i = 0, \quad \alpha_i \in [0, C]. \quad (7.36)$$

The kernel $k(x_i, x_j)$ is a function that takes two input vectors and computes a scalar product, $k(x, y) = \phi(x)^T \phi(y)$. This is popularly known as the “kernel” trick and defines the type of separating hyperplane used for classification. The kernel trick is very useful as it allows classification with non-linear hyperplanes. It allows mapping of a d -dimensional input vector into a higher dimensional H (Hilbert) feature-space using the basis function Φ , such that $x_i \rightarrow \Phi(x_i)$. The sufficient conditions necessary to guarantee the existence of the feature space are given by Mercer’s Theorem [196].

Again, after solving for the support vectors α , classification is performed using the following schemes $f(x) = \text{sign} \left(\sum_i \alpha_i y_i k(x_i, x_j) + b \right)$. There are many kernel functions that can be used, and in this work we implement the linear, quadratic, polynomial, and radial basis function (RBF) kernels. The linear kernel is trivial. The quadratic, polynomial, and RBF kernels are listed in order below.

$$k(x, y) = (x_1 y_1 + x_2 y_2)^2 \quad (7.37)$$

$$k(x, y) = (x^T y + 1)^p \quad (7.38)$$

$$k(x, y) = \exp \left(-\frac{1}{2\sigma^2} \|x - y\|^2 \right) \quad (7.39)$$

7.2.5 Cross validation and quantification of classification accuracy

In general, classification algorithms seek to determine a boundary that best separates two (or more) distinct groups of data. In this work we consider a two-class problem, where the classes were affected or not affected with RA. There are two steps to this process: (1) training and (2) testing the algorithms. In the training phase, the algorithm determines the decision boundary that best separates the training data into its two classes. In the testing phase, the ability of the algorithm to accurately classify an “unseen” data point is evaluated (this is the only way to infer how well the classification algorithm will perform on future data).

We remove any bias that may be introduced from treating each imaged finger as an independent sample by employing a modified version of the leave one-out cross-validation procedure (LOOCV) to train and test. In contrast to the standard LOOCV procedure, where one sample (finger) is used for testing while the remaining samples are used for training, we leave out all samples (fingers) belonging to one single subject (three for subjects with RA and six for subjects without RA). The remaining samples are used for training the algorithm. In the testing phase, each of the testing samples are classified as TP , TN , FP , or FN . This process is repeated for each of the 53 distinct subjects (each repetition is called an iteration).

The overall performance of the algorithm is computed by summing the TP , TN , FP , and FN values computed from each of the 53 LOOCV iterations. From these results, the sensitivity ($Se = \frac{TP}{TP+FN}$) and specificity ($Sp = \frac{TN}{TN+FP}$) values are computed for each feature combination and classification algorithm. We compute confidence intervals (CI) for Se and Sp that take into account the effective sample size (as discussed in Section 6.4.2). We report the Se and Sp values and their corresponding 95% confidence intervals (lower and upper bounds); that is, the interval within which we are confident the true value of Se and Sp lie (with 95% confidence).

7.2.6 Selection of best features

In our work to date, our strategy for finding the combination of features that yield the best results (Se and Sp) in the classification of RA has been to evaluate the ability to diagnose with all possible

feature combinations [9, 140, 173]. The same analysis cannot be performed for large numbers of features as the number of combinations is $2^n - 1 - n$, where n is the number of features. In this work $n = 594$, and the number of possible combinations is astronomical.

We overcome this problem in two steps. (1) We reduce the dimensionality l of the feature space by considering only the 30 features with the largest Youden index ($Se + Sp - 1$) computing using ROC curve analysis. (2) We employ an optimization algorithm to determine the subset of these 30 features that yield optimal or near-optimal classification results. The algorithm does not test all possible feature combinations, instead, it samples only a subset of these combinations and still achieves high classification accuracy.

We employ an evolution strategy algorithm generally referred to as $(1, \lambda)$ -single-parent evolution strategy (ES) or greedy feature selection rule. It is an optimization technique based on ideas of adaptation and evolution [189, 198]. The ES algorithm determines which multi-dimensional set of features achieves optimal (or near-optimal) Se and Sp .

Beginning with a set of parent features (p^k), the algorithm has two steps: (1) mutation and (2) selection. In the mutation step, a total of M mutants, denoted λ^k , are generated to compete with their parents p^k . The mutation occurs in three steps: (1) M_a new feature combinations are generated by adding a new feature to p^k ; (2) M_r new features are generated by replacing an existing features in p^k with one of the remaining features; (3) M_d new feature combinations are generated by dropping an existing feature in p^k . Thus, the total number of mutants M ($M_a + M_r + M_d$) in generation k are obtained by adding, dropping, or replacing one feature from parent features p^k .

In the selection step, new parents p^{k+1} are selected from the current set of parents p^k and their λ^k mutants. The selection of the new parent features p^{k+1} is made by selecting the feature combination (with dimensionality d) that yields the largest augmented Youden index, defined as

$$Y^*(Se, Sp) = Se + Sp + \alpha L_{Se} + \beta L_{Sp} - \delta d - 1, \quad (7.40)$$

where L_{Se} and L_{Sp} are the lower bounds of the confidence intervals for Se and Sp (i.e. CI_{Se}^α and CI_{Sp}^α from Chapter 6), respectively. The scaling factors α , β , and δ control the contribution of the lower bound values (L_{Se} and L_{Sp}) and dimensionality (d) of the selected feature combination on Y^* , and are all set to 0.001. In this way, feature combinations with higher lower bounds and low dimensionality are preferred.

The mutation and selection operator is applied in a loop and an iteration of the loop is called a generation, denoted by k . We begin the process by specifying the initial set of parents p^0 . The sequence of generations continues until we are unable to improve Y^* , that is $Y^{*k} = Y^{*k+1}$. The algorithm finds the feature combination with the highest Se and Sp values, favoring combinations with higher L_{Se} and L_{Sp} , and lower dimensionality d . The process is summarized as follows:

1. A feature combination (may be a single feature) is chosen as the parents p^k of the current generation k .
2. All M possible λ^k mutants of generation k are obtained by adding (M_a), dropping (M_d), or replacing (M_r) one feature from the parent combination p^k ($M = M_a + M_r + M_d$).
3. Of all λ^k mutants and their p^k parents, the combination with the largest Y^* becomes the parents of the next generation (p^{k+1}).
4. Set $k \leftarrow k + 1$ and repeat this process until the objective function Y^* does not improve (i.e. $Y^{*k} = Y^{*k+1}$).

This procedure is formulated as an optimization problem, where the objective function, ϕ_λ^k , is defined as the winning Y^* of generation k (Fig. 7.4(a)). In this sense, we seek to maximize ϕ by selecting the feature combination that maximizes Y^* at each generation and define it as:

$$\phi_\lambda^k \rightarrow \max_i (Y_\lambda^*)^k = \max_i \left(Se^{k,i} + Sp^{k,i} + \alpha L_{Se}^{k,i} + \beta L_{Sp}^{k,i} - \delta d^{k,i} - 1 \right), \quad (7.41)$$

where the index i refers to the i^{th} feature combination in generation k . $Se^{k,i}$ and $Sp^{k,i}$ are the Se and Sp from the LOOCV procedure using the i^{th} feature combination of the k^{th} generation. Similarly,

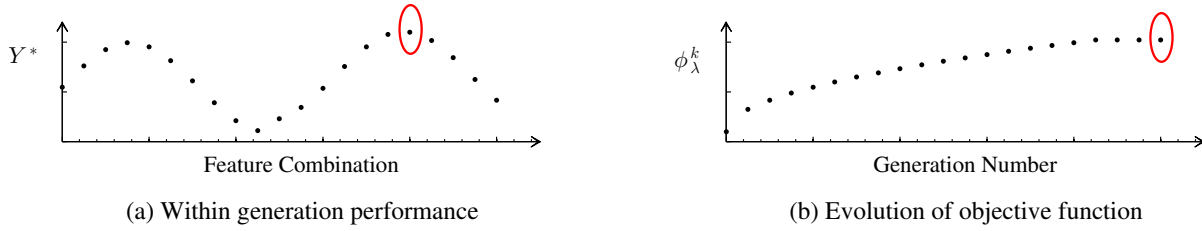


Figure 7.4: (a) Sample within-generation values of the augmented Youden index Y^* for all possible feature combinations (mutants). (b) Sample evolution of objective function over multiple generations.

$L_{Se}^{k,i}$ and $L_{Sp}^{k,i}$ are the lower bounds of the 95.0% confidence intervals for $Se^{k,i}$ and $Sp^{k,i}$, while $d^{k,i}$ is the dimensionality of the corresponding feature combination. The algorithm guarantees that $\phi_{\lambda}^{k+1} > \phi_{\lambda}^k$ until convergence, where the solution converges to a near optimal combination (a local maximum) that maximizes ϕ_{λ} (Fig. 7.4(b)). However, Se and Sp are not constrained to always increase.

To identify the appropriate parents for the first generation (p^0), we first perform LOOCV using all possible two-dimensional feature combinations and compute Y^* for each combination. Feature pairs that yield the five highest Y^* values are selected as first generation parents p^0 . Thus, five distinct optimization runs are executed for each combination of algorithm type (KNN, DA, SOM, SVM) and algorithm parameters (number of neighbors, discriminate type, SVM kernel, etc.), where the initial set of features p^0 is different for each run.

7.3 Results

We start by showing three examples of typical decision boundaries. Figure 7.5 shows these boundaries for three different classification algorithms applied to a data set consisting of two features. The two features are the minimum and mean value of μ'_s in images of all healthy subjects (blue dots) and subjects with RA (red dots). The classification algorithms that produce the decision boundaries are LDA, QDA, and SVM. In the case of LDA and QDA, all the data points influence placement of the boundary, while the support vectors determine the SVM boundary only (red circles with white filling). Here, 34 support vectors ($< 16\%$ of the data) were identified, 20 from

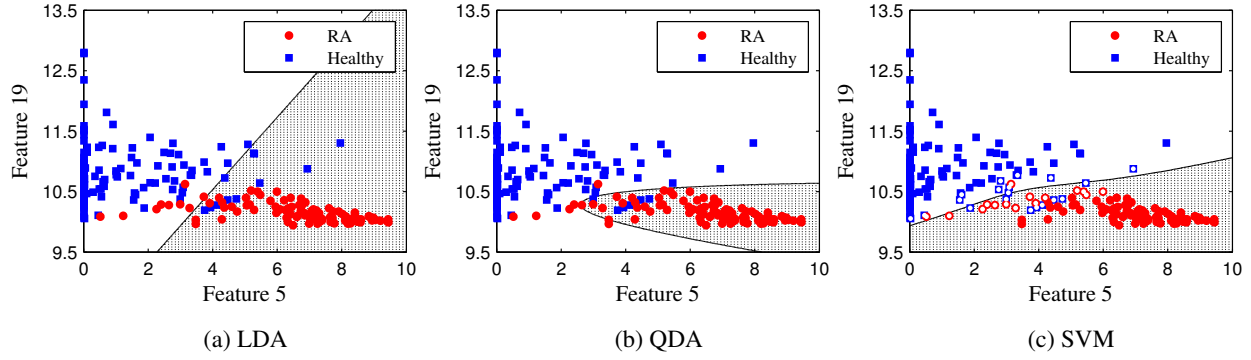


Figure 7.5: (a) LDA, (b) QDA, and (c) SVM decision boundaries separating affected from healthy data using image features 5 and 19. Support vector data points are identified by dots with white filling.

healthy subjects and 14 from subjects with RA. In these three examples the entire data is used to train (i.e. no cross-validation).

From these plots we determine TN , TP , FN , and FP values. In these example cases we find that LDA achieves 91% Se and 91% Sp ($FP = 11$, $FN = 9$), QDA results in 93% Se and 95% Sp ($FP = 6$, $FN = 7$), and SVM achieves 98% Se and 93% Sp ($FP = 9$, $FN = 2$). A similar analysis is performed for all other feature combinations identified by the feature selection algorithm, however, in those cases we do perform extensive cross-validation as described in Section 7.2.5.

Results from the evolution algorithm are summarized for each classification method in subsections 7.3.1-7.3.4. For brevity, we report only the results obtained with three distinct seed parents used in the optimization algorithm (out of a possible five). The results of non-reported optimization runs are well within the trend presented by the reported cases.

In general, the optimization algorithm always converges to a combination of two to six features in one to six generations and always improves on the initial Se and Sp by 5.0% to 15.0% for all classification methods (Fig. 7.6). The optimal set of features typically includes basic statistical features derived from the raw reconstruction data, basic features from projections of the three-dimensional data set, as well as coefficients from the Fourier transform of the data (refer to Table 7.1 for definitions). The algorithm fails to improve on the original set of feature only twice

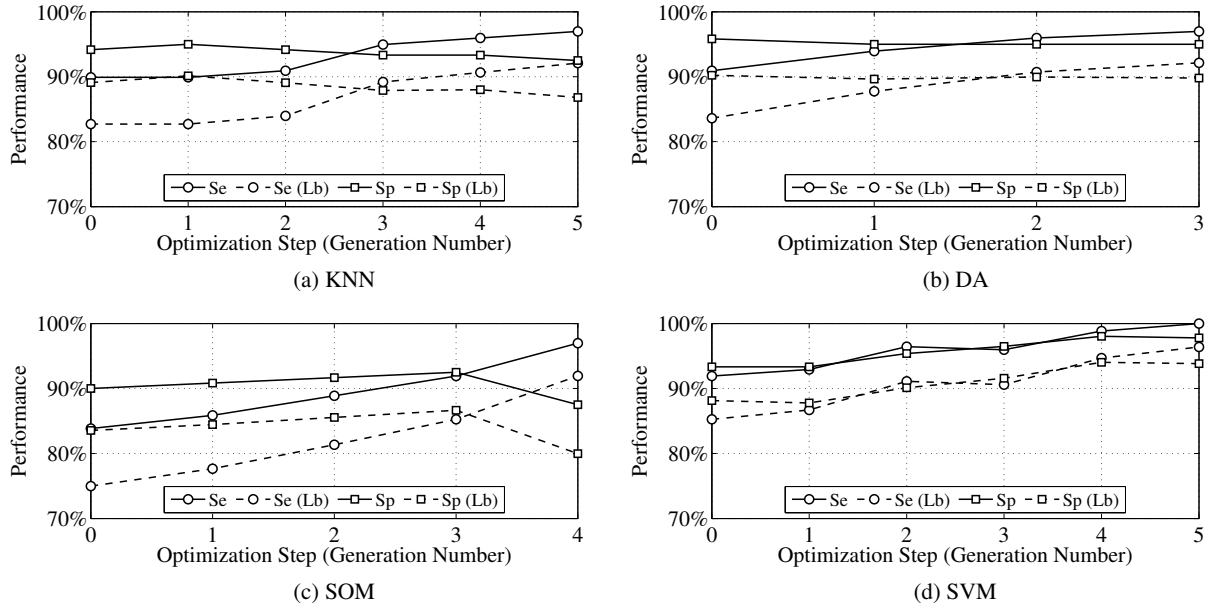


Figure 7.6: Sample Se and Sp evolution paths obtained with KNN, DA, SOM, and SVM classifiers. The lower bounds (Lb) for the 95% confidence interval are shown for Se and Sp , respectively. The upper bound in these cases is 100.0%.

(KNN, 1 neighbor, runs numbers 1 and 3).

We report the Se and Sp values to which the algorithm converges, the initial and final feature combinations, and the lower and upper bounds of the 95.0% confidence interval for Se and Sp. The following convention is used: $Se (L_{Se}, U_{Se})$, where L_{Se} and U_{Se} are the lower (L) and upper (U) bounds around the computed Se value within which we are 95.0% confident that the true value of Se lies. This interval is computed using the effective sample size (ESS) and intra-class correlation coefficient (ICC) as presented in Chapter 6, and corrects for “correlation” between fingers from the same subject. The ESS and ICC values used throughout the analysis are presented in Table 7.2.

Table 7.2: ICC and ESS values of the features listed in Table 7.1.

Feature Number	Affected		Healthy		Feature Number		Affected		Healthy	
	ICC	ESS	ICC	ESS			ICC	ESS	ICC	ESS
1	0.40	54.8	0.31	47.0	16		0.41	54.5	0.16	66.1
2	0.46	51.5	0.33	45.4	17		0.38	56.2	0.15	68.4
3	0.41	54.2	0.20	60.6	18		0.31	61.0	0.16	67.1
4	0.44	52.6	0.31	47.4	19		0.58	46.0	0.28	49.8
5	0.34	58.6	0.18	62.5	20		0.38	55.9	0.14	70.7
6	0.39	55.6	0.14	70.1	21		0.37	56.7	0.18	62.7
7	0.38	56.2	0.14	70.3	22		0.40	55.1	0.12	75.5
8	0.40	55.2	0.17	64.6	23		0.40	55.0	0.12	76.0
9	0.38	56.2	0.12	74.8	24		0.39	55.5	0.12	74.2
10	0.38	56.4	0.18	63.2	25		0.37	57.1	0.15	67.9
11	0.40	54.8	0.16	67.0	26		0.42	53.9	0.17	65.1
12	0.39	55.7	0.14	70.7	27		0.40	54.8	0.11	78.0
13	0.45	52.3	0.14	70.5	28		0.40	54.7	0.11	77.7
14	0.38	56.1	0.14	70.1	29		0.42	53.6	0.20	60.7
15	0.35	58.1	0.14	69.6	30		0.54	47.5	0.29	49.2

7.3.1 KNN

Classification with KNN is performed using seven distinct numbers of “neighbors” to ensure the general performance of the algorithm is captured ($k = 1, 3, 5, 7, 11, 15, 21$). For each k , five optimization runs, each with distinct original parent features, are executed, resulting in $7 \times 5 = 35$ distinct executions of the optimization algorithm. The best optimization runs for $k = 1, 5, 11$ are summarized in Table 7.3. Results from the remaining optimization runs are well within the trend represented by these sample results. The largest Y^* value is achieved using $k = 5$, with $Se = 96.0\%(90.7\%, 100.0\%)$ and $Sp = 94.2\%(89.0\%, 100.0\%)$. All five distinct runs with $k = 5$ converge to the same final set of features, $\{3, 8, 15, 19\}$, independent of the initial set of features. Using too few (i.e. $k = 1$) or too many ($k > 11$) neighbors results in lower Se and Sp.

Table 7.3: Classification results with the KNN algorithm using 1, 5, and 11 nearest neighbors.

Neighbors (k)	Sensitivity (95% CI)	Specificity (95% CI)	Initial Combination	Final Combination	Optimization Steps
1	93.9% (88.0%, 100.0%)	93.3% (88.0%, 100.0%)	6, 15	6, 15	1
	93.9% (87.9%, 100.0%)	92.5% (86.8%, 100.0%)	8, 19	8, 14, 19	2
	93.9% (88.1%, 100.0%)	91.7% (85.3%, 100.0%)	5, 19	5, 19	1
5	96.0% (90.7%, 100.0%)	94.2% (89.0%, 100.0%)	15, 19	3, 8, 15, 19	4
	96.0% (90.7%, 100.0%)	94.2% (89.0%, 100.0%)	6, 18	3, 8, 15, 19	5
	96.0% (90.7%, 100.0%)	94.2% (89.0%, 100.0%)	5, 19	3, 8, 15, 19	4
11	97.0% (92.1%, 100.0%)	92.5% (86.8%, 100.0%)	3, 10	7, 8, 19, 21	6
	97.0% (92.1%, 100.0%)	92.5% (86.8%, 100.0%)	3, 12	7, 8, 19, 21	6
	97.0% (92.1%, 100.0%)	92.5% (86.8%, 100.0%)	3, 6	7, 8, 19, 21	5

7.3.2 LDA and QDA

The three best classification results obtained with LDA and QDA algorithms are shown in Table 7.4. Classification with QDA is marginally better than classification with LDA. Classification with QDA converges to optimal features $\{5, 15, 19\}$ with $Se = 97.0\%(92.1\%, 100.0\%)$ and $Sp = 95.0\%(89.8\%, 100.0\%)$. Classification with LDA converges to features $\{5, 9, 15, 19\}$ with $Se = 97.0\%(92.1\%, 100.0\%)$ and $Sp = 93.3\%(87.8\%, 100.0\%)$. QDA achieves higher Se and Sp, and converges to a lower dimensional optimal feature combination compared to LDA. Both methods consistently select the same set of optimal features independent of the initial feature set, although LDA selects one additional feature compared to QDA (feature number 9).

Table 7.4: Classification results with the DA algorithm using linear and quadratic functions.

Discriminate Type	Sensitivity (95% CI)	Specificity (95% CI)	Initial Combination	Final Combination	Optimization Steps
Linear	97.0% (92.1%, 100.0%)	93.3% (87.8%, 100.0%)	5, 12	5, 9, 15, 19	4
	97.0% (92.1%, 100.0%)	93.3% (87.8%, 100.0%)	5, 19	5, 9, 15, 19	3
	97.0% (92.1%, 100.0%)	93.3% (87.8%, 100.0%)	8, 19	5, 9, 15, 19	4
Quadratic	97.0% (92.1%, 100.0%)	95.0% (89.8%, 100.0%)	5, 30	5, 15, 19	4
	97.0% (92.1%, 100.0%)	95.0% (89.8%, 100.0%)	5, 6	5, 15, 19	3
	97.0% (92.1%, 100.0%)	95.0% (89.8%, 100.0%)	8, 19	5, 15, 19	3

7.3.3 SOM

For each set of first generation parents, the optimization algorithm is executed for each possible combination of neurons ($n = 9, 16, 25$) and learning rates ($l = 0.01, 0.1, 1.0$), for a total of $9 \times 5 = 45$ optimization runs. Results from classification with SOM are summarized for $n = 9, 16$ and $l = 1.0$ in Table 7.5. While the number of neurons n did have an impact on the classification accuracy, we find that the learning rate l does not make a significant difference; for this reason only results that demonstrate dependence on n are shown.

Classification with SOMs leads to different optimal features for each run, meaning that the optimal feature combination is dependent on the initial set of features. The best classification results are $Se = 97.0\%(92.1\%, 100.0\%)$ and $Sp = 91.7\%(85.7\%, 100.0\%)$, with $n = 9$ and $l = 1.0$. The performance of SOMs with $n = 16$ is similar, however, performance significantly decreases with $n = 25$, suggesting that using 25 neurons results in over-fitting the data.

Table 7.5: Classification results with the SOM algorithm using neurons $n = 9, 16$, respectively ($l = 1.0$).

Neurons (n)	Sensitivity (95% CI)	Specificity (95% CI)	Initial Combination	Final Combination	Optimization Steps
9	89.9% (82.7%, 100.0%)	92.5% (86.9%, 100.0%)	6, 10	10, 11	2
	93.9% (88.1%, 100.0%)	90.8% (84.7%, 100.0%)	15, 19	15, 19	1
	97.0% (92.1%, 100.0%)	91.7% (85.7%, 100.0%)	19, 24	8, 17, 19	5
16	90.9% (84.0%, 100.0%)	94.2% (89.1%, 100.0%)	3, 7	3, 12	3
	94.9% (89.3%, 100.0%)	89.2% (82.7%, 100.0%)	6, 14	6, 11	2
	90.9% (84.0%, 100.0%)	94.2% (89.1%, 100.0%)	6, 18	6, 14	2

7.3.4 SVM

Classification with SVM is performed using linear, quadratic, polynomial, and RBF kernels. Classification with the polynomial kernel is explored with polynomials of degree 3, 4, 5, 6, and 7. The RBF kernel was explored by varying σ (0.1, 0.5, 1.0, 2.0, 5.0, 10.0). A total of five distinct classification runs (each with distinct features as seeds) are performed for each combination of kernel method and kernel parameter for a total of $5 + 5 + 5 \times 5 + 5 \times 6 = 65$ distinct runs.

Results from classification with linear, quadratic, and polynomial kernels are summarized in Table 7.6. Linear, quadratic, and low order polynomial kernels can separate the data well. The classification accuracy of polynomials of higher order ($p \geq 6$) is lower compared to low order polynomials. This is expected, as it is well known that higher order polynomials can severely overfit the data, resulting in poor cross-validation results. Marginally lower accuracy is obtained with the RBF kernel; $\sigma = 3$ providing the best results (omitted for brevity).

The optimal combination always converges to the same set of features, independent of initial features, and only depends on the kernel type. The largest Y^* value is achieved using a polynomial of order 3, with $Se = 100\%(96.4\%, 100.0\%)$ and $Sp = 97.8\%(93.8\%, 100.0\%)$, and converges to optimal features $\{4, 5, 6, 12, 15, 30\}$. Feature 5 is selected as an optimal classifier by all kernels, while features 15 and 19 are selected as optimal features by three different kernels.

Table 7.6: Classification results with the SVM algorithm using linear, quadratic, and polynomial kernels.

SVM Kernel	Sensitivity (95% CI)	Specificity (95% CI)	Initial Combination	Final Combination	Optimization Steps
Linear	97.0% (92.1%, 100.0%)	93.3% (87.8%, 100.0%)	5, 12	5, 9, 15, 19	4
	97.0% (92.1%, 100.0%)	93.3% (87.8%, 100.0%)	5, 19	5, 9, 15, 19	3
	97.0% (92.1%, 100.0%)	93.3% (87.8%, 100.0%)	8, 19	5, 9, 15, 19	4
Quadratic	97.0% (92.1%, 100.0%)	95.0% (89.8%, 100.0%)	5, 30	5, 15, 19	4
	97.0% (92.1%, 100.0%)	95.0% (89.8%, 100.0%)	5, 6	5, 15, 19	3
	97.0% (92.1%, 100.0%)	95.0% (89.8%, 100.0%)	8, 19	5, 15, 19	3
Polynomial Order 3	100.0% (96.4%, 100.0%)	97.8% (93.8%, 100.0%)	5, 30	4, 5, 6, 12, 15, 30	5
	100.0% (96.4%, 100.0%)	97.8% (93.8%, 100.0%)	5, 6	4, 5, 6, 12, 15, 30	5
	100.0% (96.4%, 100.0%)	97.8% (93.8%, 100.0%)	6, 27	4, 5, 6, 12, 15, 30	6
Polynomial Order 4	97.0% (91.7%, 100.0%)	94.2% (88.0%, 100.0%)	5, 6	5, 30	2
	97.0% (91.7%, 100.0%)	94.2% (88.0%, 100.0%)	6, 29	5, 30	3
	97.0% (91.7%, 100.0%)	94.2% (88.0%, 100.0%)	8, 19	5, 30	3
Polynomial Order 5	97.0% (92.2%, 100.0%)	93.3% (87.5%, 100.0%)	5, 30	5, 19	2
	97.0% (92.2%, 100.0%)	93.3% (87.5%, 100.0%)	8, 19	5, 19	2
	97.0% (92.2%, 100.0%)	93.3% (87.5%, 100.0%)	5, 19	5, 19	1

7.3.5 Best features

The frequency with which each of the original 30 features appears as an optimal classifier is presented in Fig. 7.7. Features 8, 15, and 19 are chosen as optimal features 73.0%, 40.0%, and 73.0% of the time across all KNN optimization runs (Fig. 7.7(a)). Features 5 (100.0%), 15 (80.0%), and 19 (80.0%) are chosen as optimal features most often across all DA iterations (Fig. 7.7(b)). Similarly, features 3 (20.0%), 6 (20.0%), 8 (33.0%) and 19 (33.0%) are chosen as optimal features most often across all SOM iterations (Fig. 7.7(c)). Finally, features 5 (88.0%), 6 (60.0%), 19 (48.0%), and 30 (48.0%) are chosen as optimal features most often across all SVM iterations (Fig. 7.7(d)).

Feature 5 is chosen most often by DA (100.0%) and SVM (88.0%), while feature 19 is chosen most often by KNN (73.0%), and SOM ((33.0%). Feature 19 occurs as an optimal feature at least 20.0% of the time in all classification methods. Feature 15 appears as optimal features over 20.0% of the time for three classification methods. Features 3, 5, 6, and 8 appear as an optimal feature over 20.0% of the time for two classification methods. Finally, features 7, 11, 21, and 30 are chosen at least 20.0% of the time by one classification method.

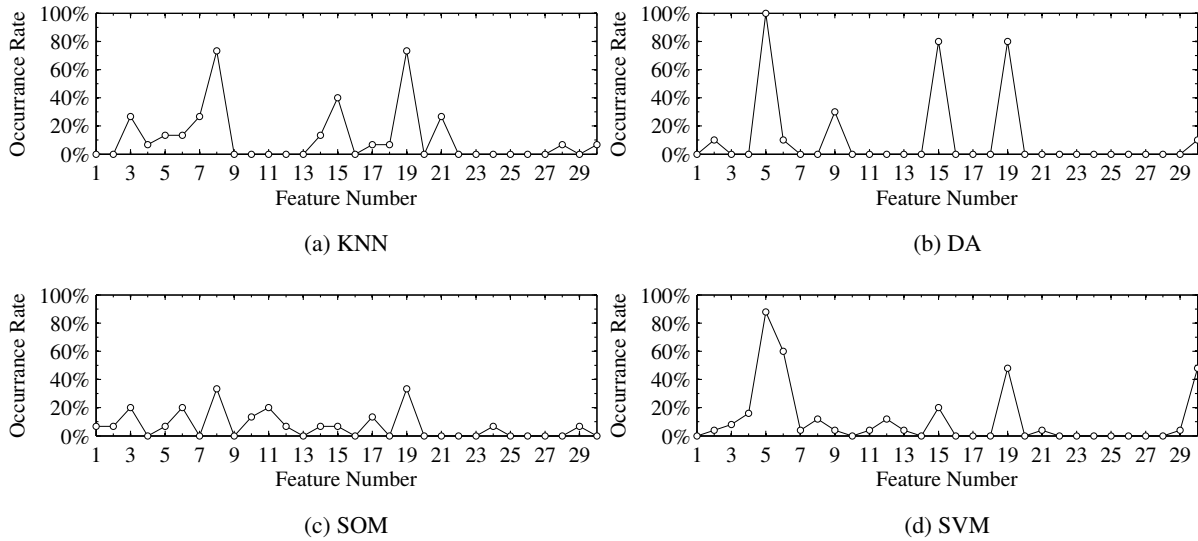


Figure 7.7: Frequency with which all features appear as optimal classifiers using KNN, DA, SOM, and SVM.

7.4 Discussion

In Chapters 6 and 7 we present a general framework for the application of computer-aided diagnosis (CAD) techniques to diffuse optical tomography (DOT). In Chapter 6 we focus on feature extraction methods, while in Chapter 7, image classification with machine learning techniques take center stage. As a specific example, the framework is applied to the classification of frequency-domain (FD) DOT images of 219 PIP joints. The goal is to classify each PIP joint as affected or not affected with RA based on the analysis of single image features (Chapter 6), and combinations of multiple image features (Chapter 7). For image classification with multiple features, we compare the performance of five different algorithms, including KNN, LDA, QDA, SOM, and SVM. Given the large number of possible permutations of the 594 features we extract, it is necessary to implement a feature-selection algorithm to determine the subset of image features that achieves the highest Se and Sp in combination with each of the five classification algorithms. Results are validated through extensive cross-validation, where the ground truth was the clinical diagnosis made according to the ACR revised criteria.

The procedure for training and testing the algorithms is a two step process using a modified version of the leave-one-out cross validation procedure: (1) the first phase uses data from $N - 1$ subjects to train the classifier, then (2) retrospective classification is performed on the data from the remaining (one) subject (testing phase). This procedure is repeated so that data from each subject (three images for each subject with RA and six images for each subject without RA) is left-out one time each.

We find that all five classification algorithms achieve clinically relevant sensitivities and specificities above 90%. The best combination of sensitivity and specificity is obtained with SVM using a polynomial kernel of degree 3. The optimal features corresponding to these results are $\{4, 5, 6, 12, 15, 30\}$. For this case, the sensitivity is 100% with a 95% confidence interval of (96.4%, 100.0%). The specificity is 97.8% with a 95% confidence interval of (93.8%, 100.0%).

Features 4 and 5 capture the range and mean μ_a and μ'_s values in the PIP joint, while features 6, 12, 15, and 30 all capture the variation of μ'_s across the joint. Feature 4 corresponds to the ratio

of maximum μ_a and minimum μ_a values in the transverse slice across the middle of the PIP joint ($F05:GT:a$); feature 5 is the minimum value of the μ'_s unstructured reconstruction ($F02:UV:s$); feature 6 is the variance value of the unstructured reconstruction μ'_s data ($F04:UV:s$); feature 12 is the mean value of the variation between all μ'_s sagittal slices ($F03:VS:s$); feature 15 is the variance value of the variation between all μ'_s coronal slices ($F04:VC:s$); feature 30 is the absolute value of the sixth coefficient of the two-dimensional-FFT of the central sagittal slices of μ'_s ($F16:VC:s$).

Features that most often achieve accurate classification are associated with global absolute values of the absorption and scattering data and their spatial variation near the PIP joint. We see evidence that features that quantify spatial variation across PIP joints are smaller for subjects with RA compared to healthy subjects. This is in concordance with our earlier findings and the images shown in Fig. 6.2 in Chapter 6. The synovial fluid as well as the surrounding tissue experience changes in optical properties in subjects with RA [135, 136]. The inflammatory process starts in the synovium, leading to changes in cell and tissue structure. Cell proliferation can be observed and the appearance of the synovial fluid changes from a clear, yellowish substance to a turbid, gray-yellowish substance. The number of leukocytes per mL increases from 100 to 200 in healthy conditions to 1,000 to 100,000 during stages 1 and 2 of the disease. Leukocytes have a diameter of approximately 7-20 μm and therefore have an effect on the scattering coefficient. Furthermore, the protein content in the synovial fluid approximately triples from 10 to 20 g/L to 30 to 60 g/L [132, 133]. In addition, neovascularization in surrounding tissue has been related to synovitis [199], which leads to an increase in the absorption coefficient.

Overall, the net effect of these changes is an increase in absorption and scattering in the finger joint affected by RA, resulting in an optical profile similar to the physiology surrounding the joint. Thus the spatial variation of optical properties decreases in subjects with RA. While similar results have been described before, it appears that our CAD algorithms can extract these features with higher accuracy and produce sensitivities and specificities at levels not reported before. All five classification algorithms achieve clinically relevant sensitivities and specificities above 90.0%, some even beyond 97.0%. The computed 95.0% confidence intervals for our results all offer further

validation that our results are robust. These observations warrant further multi-center prospective clinical studies.

One particular focus of these trials should further investigate an intriguing finding in our current study. We observed that joints of subjects with RA not yet affected by RA, as determined by MRI and US scans, were optically much more similar to joints affected by RA than joints of healthy volunteers. Indeed, the “unaffected” joints of subjects with RA exhibit statistically significant different optical features from healthy joints, yet they are optically not statistically different from affected joints. A longitudinal study will be necessary to determine if these joints eventually become affected in accordance with the ACR criteria and through MRI and US imaging. If that turns out to indeed be the case, one could prove that DOT may be used to detect changes earlier than any other imaging methods.

In addition to diagnosis of RA, we expect the framework established in this work (CAD with DOT) can be deployed in the diagnosis and monitoring of other diseases, including breast cancer and peripheral artery disease.

Outlook

If the results reported in this study are confirmed in multi-center prospective clinical trials, optical tomographic imaging could play a substantial role in clinical management of RA. DOT imaging has several advantages over MRI as well as US imaging. While MR imaging can provide anatomical features in great detail, it is not typically used for the monitoring of RA. The cost related to MR imaging are relatively high and the use of gadolinium as contrast agents is contraindicated in several cases. For example it is well known that gadolinium increases the risk of various kidney diseases. This is a particularly problematic for RA patients, since many of the drugs used to treat RA patients can have serious nephrotoxic effects by themselves [127, 128, 200]. Karie et al. recently reported that nearly half of RA patients who participated in their study presented some kidney disease.

Compared to US, optical measurement can be performed contact free. This appears to be

an advantage given the increased sensitivity to touch of joints affected by RA and the resulting discomfort experienced by patients. Furthermore, the sensitivities and specificities reported in our study compare favorably to a recent finding by Freeston et al. [201], who found a sensitivity of 0.71 and specificity of 0.82 using double ultrasound in a study involving 50 patients with RA. Furthermore, it should be noted that in current day-to-day clinical practice, medical imaging techniques are only used for confirmation of the clinical findings. Neither MRI nor US have been elevated to standard of care by the American College of Rheumatology (ACR) or the European League Against Rheumatism (EULAR) [38]. It remains to be seen if DOT can breach that gap. But given the low cost, relatively high sensitivity and specificity, as well as the contact and contrast-agent-free measurement approach, it appears that DOT could become a valuable tool in the evaluation of RA.

However, the computation time required to obtain the absorption and scattering images of a single joint is over 180 minutes. Reconstruction can take significantly longer if we use the “recommended” set of discrete ordinates ($> S_{12}$) and a sufficiently dense numerical mesh that minimizes numerical error. For practical reasons, up to this point, we have used a low level set of discrete ordinates (S_4) and relatively coarse numerical mesh. The risk in using such ordinates and mesh is the potential for numerical error to corrupt the absorption and scattering images.

Thus, we now turn our attention to improving the reconstruction process, given that we feel confident in the ability to accurately classify FD-DOT images of PIP joints as affected or not affected by RA (and indeed this can be applied to any other disease detectable by DOT). In Chapters 8 and 9 we present work that focuses on using an efficient reconstruction algorithm based on the SP_N light propagation model to reconstruct absorption and scattering from FD-DOT measurements. After presenting the theoretical and numerical steps necessary to implement the algorithm, we use the novel algorithm to reconstruct absorption and scattering in the PIP joints of the clinical data presented in Chapter 5. Subsequently, in Chapter 10 we apply the feature extraction process of Chapter 6 and image classification steps of Chapter 7 to the SP_N -based reconstruction data.

Finite volume SP_N Model: Forward Problem

In this chapter we investigate the properties and performance attributes of the two competing models of the frequency domain SP_N equations (FD- SP_N) presented in Section 2.5.3. The SP_N model developed by Dominguez (2.22), which has only real-valued diffusion coefficients, is denoted SP_N (RD). The model proposed by Chu (2.24), which has complex-valued diffusion coefficients, is referred to as SP_N (CD). This nomenclature is used extensively throughout this chapter.

We introduce the finite volume (FV) approximation for each of the FD- SP_N models (FV-FD- SP_N). We employ a node-centered FV discretization of the computational domain [202]. The resulting system of linear algebraic equations is solved with the restarted generalized minimal residual method (GMRES), which is a well-known Krylov subspace iterative method for solving systems of linear equations and famous for its computational efficiency [203]. The algorithm is implemented using objective-oriented programming in C++.

This new approach to solving the FD- SP_N equations for the application to DOT imaging is attractive because the node-centered finite volume method (FVM) takes advantage of the beneficial properties associated with the finite-element method (FEM) and the standard FVM technique.¹ It combines the conservation properties of the FVM formulation and the geometric flexibility of the FEM approach. Furthermore, our FVM implementation is advantageous over the finite differences

¹The standard FVM technique is so-called “cell-centered” as the finite volume mesh is identical to the FEM mesh. In contrast, the node-centered FVM method defined control volumes around each node in the FEM mesh. These differences are more clearly explained in Section 8.1, where a review of the finite volume mesh is provided.

method (FDM), where the computational space is discretized with Cartesian grids, because it can be easily formulated on unstructured grids to accommodate geometries of arbitrary shape. This is a particularly important property as DOT applications often involve the imaging of complex geometries (e.g. mice) that can be most accurately modeled by unstructured grids.

To date, an algorithm for solving the SP_N equations with an FVM approach has not yet been developed. The SP_N equations have been solved in the steady state with an FDM approach by Klose et al. [1], in the frequency domain with an FEM approach by Chu et al. [73], and in the time domain with an FEM approach by Dominguez et al. [74]. The algorithms for solving the SP_N equations on unstructured grids reported to date have lacked computational efficiency.² Klose et al. achieved the desired computational speed improvements with an FDM approach, however, accurately modeling curved boundaries with structured grids remains a challenge and may have only limited success in various clinical and pre-clinical applications [66].

The FEM implementation of the SP_N equations by Dominguez et al. did not report computation time so an accurate evaluation of their algorithm's efficiency is not possible [74]. The implementation of the FD- SP_N equations by Chu et al. [73] reported results from simulations on a numerical phantom with 42,386 mesh nodes (235,869 linear tetrahedral elements), $\mu_a = 0.1 \text{ cm}^{-1}$, $\mu_s = 200.0 \text{ cm}^{-1}$, $g = 0.5$, and $n_m = 1.37$. The computation time for the forward solution of this model ($4 \times 2 \times 3 \text{ cm}^3$) was 4.2, 12.6, 31.7, and 69.7 minutes for the SP_1 , SP_3 , SP_5 , and SP_7 , respectively. These computation times can be significantly improved with a more efficient implementation. Furthermore, this algorithm by Chu et al. was implemented in MATLAB, a computational program that is known for ease of use but can suffer from slow execution time when compared to algorithms in C or C++.

In this chapter we present the necessary framework for the implementation of an algorithm that solves the FV-FD- SP_N equations. Our C++ based FV-FD- SP_N algorithm is computationally efficient, requiring low system memory and converging to acceptable solutions significantly

²We note that the reported computations were performed on a 3.0 Ghz, 64 bit Linux machine with 8 GB of physical memory. In contrast, the machine in use of this work reports CPU speeds of 2.6 GHz (running OS X 10.8 with 12 GB of physical memory).

Table 8.1: Definition of variables.

Variable	Description	Units
FVM	Finite volume method	
FEM	Finite element method	
FDM	Finite differences method	
\mathbf{O}	Imaging object	
\mathbf{V}	Computational domain (i.e. discretization of \mathbf{O})	
M	Total number of discrete mesh elements in \mathbf{V}	
m	Discrete mesh element, s.t. $\mathbf{V} = \{m : m = 1 \dots M\}$	
S_B	Number of boundary surfaces of the m th FV element	
S_I	Number of internal surfaces of the m th FV element	
S_T	Total number of finite volume surfaces in \mathbf{V}	
N	Order of SP_N model	
μ_a	Absorption coefficient	cm^{-1}
μ_s	Scattering coefficient	cm^{-1}
μ'_s	Reduced scattering coefficient	cm^{-1}
μ_{an}	n th order absorption coefficient	cm^{-1}
g	Anisotropy factor	
φ_n	n th order composite moment of the SP_N model	W cm^{-2}
ϕ_n	n th order Legendre moments of the radiance	W cm^{-2}
$\hat{\mathbf{n}}$	Normal vector pointing outwards at outer surface of \mathbf{O}	
Ω	Angular direction	
\mathbf{r}	Spatial position	
v	Speed of light in tissue	cm s^{-1}
ω	Modulation angular frequency ($2\pi f$)	s^{-1}
Q	Interior source	W cm^{-3}
S	Boundary source	$\text{W cm}^{-2}\text{sr}^{-1}$
J^+	Partial current	W cm^{-2}
J_n	Coefficients of the partial current	
i	Complex unit ($\sqrt{-1}$)	

faster than previously reported algorithms that use unstructured grids. The algorithm is capable of accurately modeling light propagation in geometries of arbitrary shape as the imaging object is discretized with unstructured grids. We validate the algorithm through numerical simulations, using the ERT model as a benchmark. In Chapter 9 we use the FV-FD-SP_N algorithm as the basis of an efficient DOT reconstruction algorithm.

A general overview of the FVM technique is presented in Section 8.1. Then, the derivation of the FV-FD-SP_N model is developed over Sections 8.2-8.3. Results from numerical simulations are presented in Section 8.4. Chapter 8 concludes with a discussion in Section 8.5.

8.1 Finite volume mesh

The imaging domain \mathbf{O} is discretized with a computational domain \mathbf{V} , composed of linked unstructured Euclidean simplex elements (Fig. 8.1). These shapes are triangles in two-dimensional space and tetrahedra in three-dimensional space. The discretized domain \mathbf{V} consists of M mesh nodes, where each mesh node m ($m \in M$) is the center of a finite volume dV (there are M distinct finite volume elements). Thus, m refers to a specific mesh node and also to a specific finite volume element as pictured in Fig. 8.1b. Finite volume element m has S_I internally facing surfaces (i.e. not on the surface of \mathbf{V}) and S_B boundary surfaces (i.e. on the surface of \mathbf{V}). This mesh is generally referred to as an FEM mesh because it is typically used “as is” when solving systems of equations with an FEM algorithm.

Without loss of generality, consider a two-dimensional object discretized by triangular elements (three dimensional objects would be discretized by tetrahedral elements). Figure 8.1 is an example of a simple two-dimensional object and its corresponding discretization into an unstructured triangular FEM mesh (red) with its corresponding unstructured FVM mesh superimposed (blue). In general, \mathbf{O} is first discretized into an unstructured triangular mesh, from which the FVM mesh is subsequently computed.

The mesh conversion process refers to finding the dual FVM mesh (blue mesh) of the original

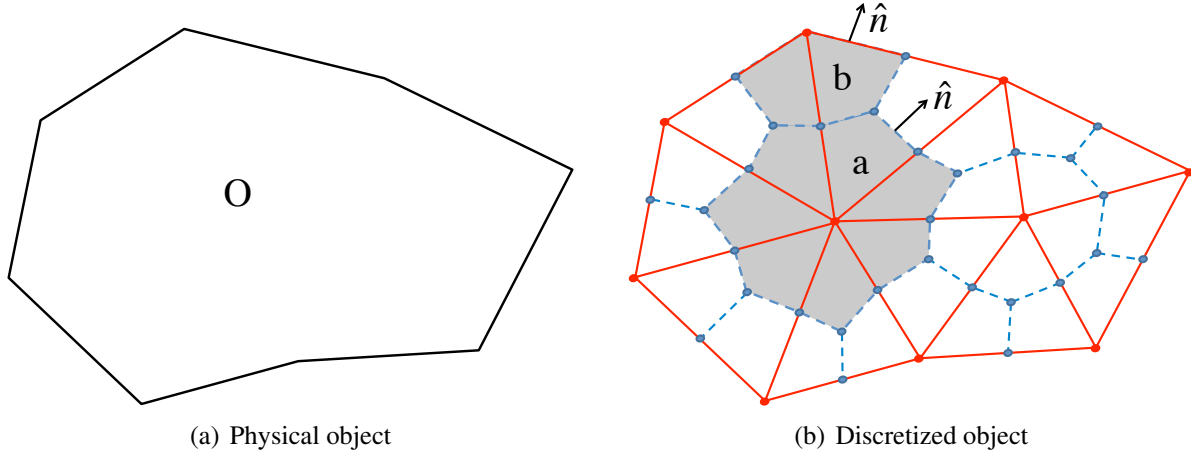


Figure 8.1: (a) An example physical object O , (b) and its corresponding discretization into triangular elements (red) and finite volume (FV) elements (blue). Each surface of the FV element has a corresponding normal vector \hat{n} .

FEM mesh (red mesh). Each node in the FEM mesh (red dots) becomes the center of an FVM element, which is surrounded with faces or edges (blue dashed lines) that are formed by the centers (blue dots) of edges, faces, and elements of the FEM mesh (Fig. 8.1b). There are two general methods for obtaining the FVM mesh from an FEM mesh, the so called “median” and “Delaunay” methods. Both methods require knowledge of the center of each triangular element in the FEM mesh and the centers of all its edges and faces (blue dots). However, each method has a specific way of defining the “center” of these geometries. The two methods are defined below.

1. Median dual method: treat the centroid of the edge, face or element as its center point (Fig. 8.2a).
2. Delaunay dual method: treat the center of the circumsphere of the edge, face or element as its center point (Fig. 8.2b).

The median dual method is used in this work. The median dual mesh method is attractive because it guarantees that the center of each FEM element is inside the simplex element, while this is not always the case with the Delaunay method. The Delaunay dual method has extra requirements on the geometry of the simplex elements. When these requirements are not met, the computed center of the simplex element may be outside the element itself, which can result in a

poorly structured FVM mesh. Figure 8.2b is an example of a triangular element whose Delaunay center is located outside the element itself.

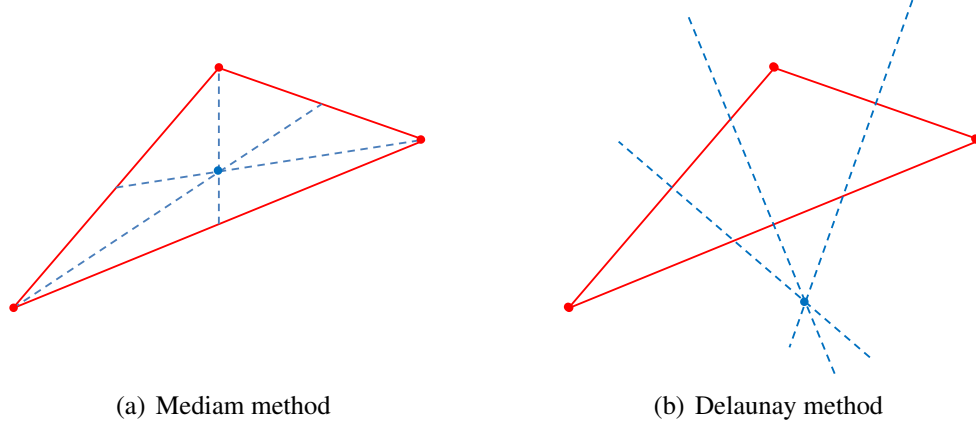


Figure 8.2: Geometric center (blue dot) computed with the (a) median and (b) Delaunay methods.

8.2 Finite volume FD-SP_N approximation

In this section we focus entirely on the FD-SP₃ model and the derivation of its finite volume formulation. Formulae for the finite volume form of other orders of the SP_N models (i.e. $N = 1, 5$, or 7) can be derived by following the same procedures presented for the SP₃ model and are omitted for brevity.

We derive the FVM approximation to the FD-SP₃ (RD) model in detail. The derivation of the FVM approximation for the FD-SP₃ (CD) model follows the same framework and is omitted for brevity. The FD-SP₃ (RD) and FD-SP₃ (CD) models are reproduced in (8.1-8.2) and (8.3-8.4), respectively.

$$-\nabla \cdot \frac{1}{3\mu_{a1}} \nabla \varphi_1 + \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 = Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_2 \quad (8.1)$$

$$-\nabla \cdot \frac{1}{7\mu_{a3}} \nabla \varphi_2 + \left(\frac{4}{9}\mu_a + \frac{5}{9}\mu_{a2} + \frac{1}{3}\frac{i\omega}{v} \right) \varphi_2 = -\frac{2}{3}Q + \frac{2}{3}\mu_a\varphi_1 \quad (8.2)$$

$$-\nabla \cdot \frac{1}{3\bar{\mu}_{a1}} \nabla \varphi_1 + \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 = Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_2 \quad (8.3)$$

$$-\nabla \cdot \frac{1}{7\bar{\mu}_{a3}} \nabla \varphi_2 + \left(\frac{4}{9}\mu_a + \frac{5}{9}\bar{\mu}_{a2} + \frac{i\omega}{v} \right) \varphi_2 = -\frac{2}{3}Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 \quad (8.4)$$

The composite moments $\varphi_1 = \varphi_1(\mathbf{r}, \omega)$ and $\varphi_2 = \varphi_2(\mathbf{r}, \omega)$ are functions of space and modulation frequency. The corresponding set of boundary equations for the FD-SP₃ (RD) and FD-SP₃ (CD) models are given by (8.5-8.6) and (8.7-8.8), respectively.

$$\begin{aligned} \left(\frac{1}{2} + A_1 \right) \varphi_1 + \left(\frac{1+B_1}{3\mu_{a1}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) &= \left(\frac{1}{8} + C_1 \right) \varphi_2 + \left(\frac{D_1}{\mu_{a3}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2) \\ &+ \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) 2|\Omega \cdot \hat{\mathbf{n}}| d\Omega, \end{aligned} \quad (8.5)$$

$$\begin{aligned} \left(\frac{7}{24} + A_2 \right) \varphi_2 + \left(\frac{1+B_2}{7\mu_{a3}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2) &= \left(\frac{1}{8} + C_2 \right) \varphi_1 + \left(\frac{D_2}{\mu_{a1}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) \\ &+ \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) (5|\Omega \cdot \hat{\mathbf{n}}|^3 - 3|\Omega \cdot \hat{\mathbf{n}}|) d\Omega. \end{aligned} \quad (8.6)$$

$$\begin{aligned} \left(\frac{1}{2} + A_1 \right) \varphi_1 + \left(\frac{1+B_1}{3\bar{\mu}_{a1}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) &= \left(\frac{1}{8} + C_1 \right) \varphi_2 + \left(\frac{D_1}{\bar{\mu}_{a3}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2) \\ &+ \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) 2|\Omega \cdot \hat{\mathbf{n}}| d\Omega, \end{aligned} \quad (8.7)$$

$$\begin{aligned} \left(\frac{7}{24} + A_2 \right) \varphi_2 + \left(\frac{1+B_2}{7\bar{\mu}_{a3}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2) &= \left(\frac{1}{8} + C_2 \right) \varphi_1 + \left(\frac{D_2}{\bar{\mu}_{a1}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) \\ &+ \int_{\Omega \cdot \hat{\mathbf{n}} < 0} S(\Omega) (5|\Omega \cdot \hat{\mathbf{n}}|^3 - 3|\Omega \cdot \hat{\mathbf{n}}|) d\Omega. \end{aligned} \quad (8.8)$$

The definitions of the partial current, $J^+(\mathbf{r}, \omega)$, at the boundary of the object for the FD-SP₃ (RD) and FD-SP₃ (CD) models are given by (8.9) and (8.10), respectively.

$$\begin{aligned}
J^+(\mathbf{r}, \omega) &= \left(\frac{1}{4} + J_0 \right) \left(\varphi_1 - \frac{2}{3} \varphi_2 \right) - \left(\frac{0.5 + J_1}{3\mu_{a1}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) \\
&\quad + \left(\frac{5}{16} + J_2 \right) \left(\frac{1}{3} \varphi_2 \right) - \left(\frac{J_3}{7\mu_{a3}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2).
\end{aligned} \tag{8.9}$$

$$\begin{aligned}
J^+(\mathbf{r}, \omega) &= \left(\frac{1}{4} + J_0 \right) \left(\varphi_1 - \frac{2}{3} \varphi_2 \right) - \left(\frac{0.5 + J_1}{3\bar{\mu}_{a1}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) \\
&\quad + \left(\frac{5}{16} + J_2 \right) \left(\frac{1}{3} \varphi_2 \right) - \left(\frac{J_3}{7\bar{\mu}_{a3}} \right) (\hat{\mathbf{n}} \cdot \nabla \varphi_2).
\end{aligned} \tag{8.10}$$

We note that the only difference between the FD-SP₃ (RD) and FD-SP₃ (CD) models is the definition of the n th order absorption coefficient. The n th order absorption coefficients for the FD-SP₃ (RD) model (8.1,8.2, 8.5,8.6,8.9) is $\mu_{an} = \mu_a + (1 - g^n) \mu_s$, while for the FD-SP₃ (CD) model (8.3,8.4,8.7,8.8,8.10) it is $\bar{\mu}_{an} = \mu_a + (1 - g^n) \mu_s + \frac{i\omega}{v}$.

The fluence, ϕ_n , is computed from the individual composite moments of the SP_N model and is given by,

$$\phi_0 = \varphi_1 - \frac{2}{3} \varphi_2, \tag{8.11}$$

$$\phi_2 = \frac{1}{3} \varphi_2. \tag{8.12}$$

8.2.1 Re-writing of the FD-SP₃ system

Our attention now shifts to the derivation of the finite volume formulation of the FD-SP₃ (RD) model.³ For simplicity, we refer to the FD-SP₃ (RD) model simply as “SP₃” in this section. For convenience, we re-write the SP₃ equations (8.1, 8.2, 8.5, 8.6) as follows:

³We remind the reader that we omit the derivation of the finite volume formulation of the FD-SP₃ (CD) model for brevity. However, the derivation is almost identical to the derivation of the finite volume formulation of the FD-SP₃ (RD) model.

$$-\nabla \cdot \mathfrak{D}_1 \nabla \varphi_1 + \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 = Q + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_2, \quad (8.13)$$

$$-\nabla \cdot \mathfrak{D}_2 \nabla \varphi_2 + \left(\frac{4}{9} \mu_a + \frac{5}{9} \mu_{a2} + \frac{1}{3} \frac{i\omega}{v} \right) \varphi_2 = -\frac{2}{3} Q + \frac{2}{3} \mu_a \varphi_1, \quad (8.14)$$

with the following set of boundary equations,

$$\alpha_1 \varphi_1 + \mathfrak{D}_1 \beta_1 (\hat{\mathbf{n}} \cdot \nabla \varphi_1) = \gamma_1 \varphi_2 + 7 \mathfrak{D}_2 \delta_1 (\hat{\mathbf{n}} \cdot \nabla \varphi_2) + S_1, \quad (8.15)$$

$$\alpha_2 \varphi_2 + \mathfrak{D}_2 \beta_2 (\hat{\mathbf{n}} \cdot \nabla \varphi_2) = \gamma_2 \varphi_1 + 3 \mathfrak{D}_1 \delta_2 (\hat{\mathbf{n}} \cdot \nabla \varphi_1) + S_2. \quad (8.16)$$

The corresponding formula for the partial current is

$$\begin{aligned} J^+ = & \left(\frac{1}{4} + J_0 \right) \left(\varphi_1 - \frac{2}{3} \varphi_2 \right) - \mathfrak{D}_1 (0.5 + J_1) (\hat{\mathbf{n}} \cdot \nabla \varphi_1) \\ & + \left(\frac{5}{16} + J_2 \right) \left(\frac{1}{3} \varphi_2 \right) - \mathfrak{D}_2 (J_3) (\hat{\mathbf{n}} \cdot \nabla \varphi_2). \end{aligned} \quad (8.17)$$

Coefficients that appear in (8.1-8.17) are defined and summarized in Tables 8.1, 8.2, and 8.3.

8.2.2 Finite volume derivation of FD-SP₃ system

Begin by integrating (8.13-8.14) and applying Stoke's theorem to convert the volume integral (defined within a finite volume element of total volume ΔV) into a surface integral along the surfaces S of the volume element, such that

$$-\int_S \mathfrak{D}_1 (\nabla \varphi_1 \cdot \hat{\mathbf{n}}) dA + \left(\mu_a + \frac{i\omega}{v} \right) \varphi_1 \Delta V = Q \Delta V + \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) \varphi_2 \Delta V, \quad (8.18)$$

$$-\int_S \mathfrak{D}_2 (\nabla \varphi_2 \cdot \hat{\mathbf{n}}) dA + \left(\frac{4}{9} \mu_a + \frac{5}{9} \mu_{a2} + \frac{1}{3} \frac{i\omega}{v} \right) \varphi_2 \Delta V = -\frac{2}{3} Q \Delta V + \frac{2}{3} \mu_a \varphi_1 \Delta V. \quad (8.19)$$

Table 8.2: Boundary and current coefficients for the SP_3 model as defined by Klose and Larsen in [1].

$$\begin{aligned}
\mathfrak{D}_1 &= \frac{1}{3\mu_{a1}} & \xi_{11} &= \frac{\alpha_1\beta_2 - 7\delta_1\gamma_2}{\beta_1\beta_2 - 21\delta_1\delta_2} \\
\mathfrak{D}_2 &= \frac{1}{7\mu_{a3}} & \xi_{12} &= \frac{7\delta_1\alpha_2 - \gamma_1\beta_2}{\beta_1\beta_2 - 21\delta_1\delta_2} \\
\alpha_1 &= \frac{1}{2} + A_1 & \xi_{13} &= \frac{\beta_2}{\beta_1\beta_2 - 21\delta_1\delta_2} \\
\alpha_2 &= \frac{7}{24} + A_2 & \xi_{14} &= \frac{7\delta_1}{\beta_1\beta_2 - 21\delta_1\delta_2} \\
\beta_1 &= 1 + B_1 & \xi_{21} &= \frac{3\delta_2\xi_{11} - \gamma_2}{\beta_2} \\
\beta_2 &= 1 + B_2 & \xi_{22} &= \frac{3\delta_2\xi_{12} + \alpha_2}{\beta_2} \\
\gamma_1 &= \frac{1}{8} + C_1 & \xi_{23} &= \frac{3\delta_2\xi_{13}}{\beta_2} \\
\gamma_2 &= \frac{1}{8} + C_2 & \xi_{24} &= \frac{3\delta_2\xi_{14} + 1}{\beta_2} \\
\delta_1 &= D_1 & \phi_0 &= \varphi_1 - \frac{2}{3}\varphi_2 \\
\delta_2 &= D_2 & \phi_1 &= \frac{1}{3}\varphi_2 \\
S_1 &= \int_{\mathbf{\Omega} \cdot \hat{\mathbf{n}} < 0} S(\mathbf{\Omega}) 2|\mathbf{\Omega} \cdot \hat{\mathbf{n}}| d\Omega \\
S_2 &= \int_{\mathbf{\Omega} \cdot \hat{\mathbf{n}} < 0} S(\mathbf{\Omega}) (5|\mathbf{\Omega} \cdot \hat{\mathbf{n}}|^3 - 3|\mathbf{\Omega} \cdot \hat{\mathbf{n}}|) d\Omega
\end{aligned}$$

The surface integrals in (8.18-8.19) are computed using numerical quadrature rules, exploiting the fact that each volume element has a finite number of surfaces (typically less than 12 surfaces). The directional derivative (boundary flux) terms ($\nabla\varphi_1 \cdot \hat{\mathbf{n}}$ and $\nabla\varphi_2 \cdot \hat{\mathbf{n}}$) are simplified by using these same properties of the finite volume element, and are written in discretized form as follows,

$$\nabla\varphi_1 \cdot \hat{\mathbf{n}} \approx \frac{[\varphi_1]_i - [\varphi_1]_m}{dr_i}, \quad (8.20)$$

$$\nabla\varphi_2 \cdot \hat{\mathbf{n}} \approx \frac{[\varphi_2]_i - [\varphi_2]_m}{dr_i}. \quad (8.21)$$

Consider an individual finite volume element $m \in V$ with S_I internal surfaces and S_B boundary surfaces (Fig. 8.1b). By definition, element m shares each of the surfaces S_I with a single neighboring finite volume element. Then, let i denote the i th neighbor of m , such that $i = 1 \dots S_I$. Then, the finite volume version (i.e. discrete formulae) of (8.13-8.14) are

$$\begin{aligned} -\sum_{i=1}^{S_I} \mathfrak{D}_1 \frac{[\varphi_1]_i - [\varphi_1]_m}{dr_i} dA_i + \left(\mu_a + \frac{i\omega}{v} \right) [\varphi_1]_m \Delta V &= [Q]_m \Delta V \\ &+ \frac{2}{3} \left(\mu_a + \frac{i\omega}{v} \right) [\varphi_2]_m \Delta V, \end{aligned} \quad (8.22)$$

$$\begin{aligned} -\sum_{i=1}^{S_I} \mathfrak{D}_2 \frac{[\varphi_2]_i - [\varphi_2]_m}{dr_i} dA_i + \left(\frac{4}{9}\mu_a + \frac{5}{9}\mu_{a2} + \frac{1}{3}\frac{i\omega}{v} \right) [\varphi_2]_m \Delta V &= -\frac{2}{3}[Q]_m \Delta V \\ &+ \frac{2}{3}\mu_a [\varphi_1]_m \Delta V. \end{aligned} \quad (8.23)$$

8.2.3 Finite volume derivation of the boundary equations

Solving boundary equations (8.15-8.16) for $(\hat{\mathbf{n}} \cdot \nabla\varphi_1)$ and $(\hat{\mathbf{n}} \cdot \nabla\varphi_2)$ yields,

$$(\hat{\mathbf{n}} \cdot \nabla \varphi_1) = \frac{1}{\mathfrak{D}_1} (-\xi_{11}\varphi_1 - \xi_{12}\varphi_2 + \xi_{13}S_1 + \xi_{14}S_2), \quad (8.24)$$

$$(\hat{\mathbf{n}} \cdot \nabla \varphi_2) = \frac{1}{\mathfrak{D}_2} (-\xi_{21}\varphi_1 - \xi_{22}\varphi_2 + \xi_{23}S_1 + \xi_{24}S_2). \quad (8.25)$$

Definitions of the coefficients ξ are summarized in Table 8.2. Equations (8.24-8.25) can be used on (8.18-8.19) to solve for the following system of discretized boundary equations.

$$\begin{aligned} \sum_{i=1}^{S_B} \xi_{11}[\varphi_1]_m dA_i + \sum_{i=1}^{S_B} \xi_{12}[\varphi_2]_m dA_i + \left(\mu_a + \frac{i\omega}{v}\right) [\varphi_1]_m \Delta V - \frac{2}{3} \left(\mu_a + \frac{i\omega}{v}\right) [\varphi_2]_m \Delta V \\ = [Q]_m \Delta V + \sum_{i=1}^{S_B} \xi_{13}[S_1]_p dA_i + \sum_{i=1}^{S_B} \xi_{14}[S_4]_p dA_i \end{aligned} \quad (8.26)$$

$$\begin{aligned} \sum_{i=1}^{S_B} \xi_{21}[\varphi_1]_m dA_i + \sum_{i=1}^{S_B} \xi_{22}[\varphi_2]_m dA_i + \left(\frac{4}{9}\mu_a + \frac{5}{9}\mu_{a2} + \frac{i\omega}{v}\right) \varphi_2 \Delta V - \frac{2}{3}\mu_a[\varphi_1]_m \Delta V \\ = -\frac{2}{3}[Q]_m \Delta V + \sum_{i=1}^{S_B} \xi_{23}[S_1]_p dA_i + \sum_{i=1}^{S_B} \xi_{24}[S_2]_p dA_i \end{aligned} \quad (8.27)$$

8.2.4 Partial current operator

The partial current, J^+ , can be discretized by using (8.24-8.25) on (8.17). This operation results in the following discretized partial current operator that acts on the fluence moments φ_n and boundary sources S_n .

$$[J^+]_p = \nu_0[\varphi_1]_p + \nu_1[\varphi_2]_p + \nu_2[S_1]_p + \nu_3[S_2]_p \quad (8.28)$$

The definition of the coefficients ν_n in (8.28) are defined in (8.29).

$$\begin{aligned}
\nu_0 &= \left[\frac{1}{4} + J_0 + \left(\frac{1}{2} + J_1 \right) \xi_{11} + J_3 \xi_{21} \right] \\
\nu_1 &= \left[-\frac{3}{48} - \frac{2}{3} J_0 + \frac{1}{3} J_2 + \left(\frac{1}{2} + J_1 \right) \xi_{12} + J_3 \xi_{22} \right] \\
\nu_2 &= - \left[\left(\frac{1}{2} + J_1 \right) \xi_{13} + J_3 \xi_{23} \right] \\
\nu_3 &= - \left[\left(\frac{1}{2} + J_1 \right) \xi_{14} + J_3 \xi_{24} \right]
\end{aligned} \tag{8.29}$$

8.3 Algebraic formulation

8.3.1 Forward model

The FD-SP₃ model given by (8.13-8.14) is described in the FVM formulation by (8.22, 8.23, 8.26, 8.27). These equations can be used to solve for the composite moments of the fluence, φ_1 and φ_2 . The fluence or Legendre moments of the radiance, ϕ_n , can be computed from φ_n . In DOT, the quantity of interest is the partial current, $J^+(\mathbf{r}, \omega)$, which is a measure of the energy that exits the medium at surface point \mathbf{r} and is measured by detectors. This partial current can be computed by applying (8.28) to φ_n .⁴

It is convenient to convert the system of equations in (8.22, 8.23, 8.26, 8.27) into an algebraic problem that is completely independent of the underlying physical problem (i.e. the computational mesh). The forward problem for the SP₃ model can be formulated into a system of algebraic equations of the form

$$Ax = b, \tag{8.30}$$

where A is the “forward model” (i.e. FV-FD-SP_N model), x is the “forward solution” (i.e. composite moments of the FV-FD-SP_N model), and b denotes the source terms for the forward problem.

⁴Here we assume source points cannot be detector points at the same time. Thus, the boundary source terms, S_1 and S_2 , do not contribute to $J^+(\mathbf{r}, \omega)$ and are ignored in (8.28).

Table 8.3: Definition of variables for the linear problem of the SP₃ model.

Variable	Description	Dimensions
A	Forward model	$\mathbb{C}^{2M \times 2M}$
b	Source vector	\mathbb{C}^{2M}
x	Solution vector (i.e. composite moments of SP _N model)	\mathbb{C}^{2M}
\mathbf{Q}	Projection operator; maps forward solution to detector readings	$\mathbb{C}^{p \times M}$
p	Number of source/detector pairs	
$\hat{\mu}_a$	Rescaled absorption coefficient	
$\hat{\mu}_{sa}$	Rescaled scattering coefficient	
l_a	Linear coefficient of the linear transformation of μ_a	
l_s	Linear coefficient of the linear transformation of μ_s	
c_a	Constant coefficient of the linear transformation of μ_a	
c_s	Constant coefficient of the linear transformation of μ_s	

This formulation is convenient because it is compact and it allows for the application of well established algebraic techniques for solving for x and differentiating the entire expression with respect to x .

Matrix A is a square matrix of size $M \times M$ for SP₁ and $2M \times 2M$ for SP₃, where M is the number of distinct mesh nodes in the numerical grid that discretizes the imaging volume \mathbf{O} . The elements of A are given by the left hand side terms of (8.22, 8.23, 8.26, 8.27) that do not contain S (boundary source) or Q (internal source). The contributions from associated boundary and internal source terms are stored in b , a vector of length M for SP₁ and $2M$ for SP₃. The elements of b are given by the terms in (8.22, 8.23, 8.26, 8.27) that contain S and Q .

The forward solution, x , is a vector of length M for SP₁ and $2M$ for SP₃. Specifically, $x = \varphi_1$ for SP₁ and $x = [\varphi_1^\top \ \varphi_2^\top]^\top$ for SP₃. The algebraic system given by (8.30) allows for the solution of φ_1 and φ_2 to be obtained simultaneously. The conjugate transpose of the complex vectors φ_1 and φ_2 are denoted by \top .

For clarity and for improved computational performance, it is helpful to rewrite the final discretized form of the SP₃ model defined by (8.22, 8.23, 8.26, 8.27). The absorption and scattering coefficients are inherently in significantly different scales ($\mu_s \gg \mu_a$), which can lead to slow convergence during the reconstruction process. This issue can be overcome by defining scaled

absorption ($\hat{\mu}_a$) and scattering ($\hat{\mu}_s$) coefficients. The scaled coefficients are

$$\begin{aligned}\hat{\mu}_a &= l_a \mu_a + c_a, \\ \hat{\mu}_s &= l_s \mu_s + c_s,\end{aligned}$$

where l_a , c_a , l_s , and c_s are computed from predefined minimum and maximum allowed values for μ_a and μ_s . For example, $\mu_a \in [0.0, 1.0] \text{ cm}^{-1}$ and $\mu_s \in [0.0, 200.0] \text{ cm}^{-1}$ (assuming the anisotropic factor $g \approx 0.9$). We note that these upper and lower bounds are for rescaling purposes only and should not be confused with constraints on the admissible values of the variables.

The n th order diffusion operator (\mathfrak{D}_n) appearing in (8.22-8.23) are defined at the face of the finite volume element and need to be replaced with corresponding nodal values (i.e. the value at the center of the finite volume element). This is achieved by replacing the n th order diffusion coefficients (\mathfrak{D}_1 and \mathfrak{D}_2) with the average of these coefficients between node m and its neighboring element i , as follows

$$\begin{aligned}\mathfrak{D}_1 &= \frac{[\mathfrak{D}_1]_m + [\mathfrak{D}_1]_i}{2}, \\ \mathfrak{D}_2 &= \frac{[\mathfrak{D}_2]_m + [\mathfrak{D}_2]_i}{2}.\end{aligned}$$

Thus, the final version of the SP_3 model can be obtained by expanding the \mathfrak{D}_1 and \mathfrak{D}_2 terms, then expanding all μ_{an} terms, and then replacing μ_a and μ_s with $\hat{\mu}_a$ and $\hat{\mu}_s$ throughout. This leads to the following expanded expressions for \mathfrak{D}_1 and \mathfrak{D}_2 :

$$\mathfrak{D}_1 = \frac{1}{2} \left[\frac{1}{3 \left(\frac{[\hat{\mu}_a]_m - c_a}{l_a} + (1-g) \frac{[\hat{\mu}_s]_m - c_s}{l_s} \right)} + \frac{1}{3 \left(\frac{[\hat{\mu}_a]_i - c_a}{l_a} + (1-g) \frac{[\hat{\mu}_s]_i - c_s}{l_s} \right)} \right], \quad (8.31)$$

$$\mathfrak{D}_2 = \frac{1}{2} \left[\frac{1}{7 \left(\frac{[\hat{\mu}_a]_m - c_a}{l_a} + (1-g^3) \frac{[\hat{\mu}_s]_m - c_s}{l_s} \right)} + \frac{1}{7 \left(\frac{[\hat{\mu}_a]_i - c_a}{l_a} + (1-g^3) \frac{[\hat{\mu}_s]_i - c_s}{l_s} \right)} \right]. \quad (8.32)$$

Matrix A is a symmetric sparse matrix, with up to $2S_T + 4M$ non-zero elements for the SP_3 system, where S_T is the total number of finite volume surfaces in a mesh that has M mesh nodes. The elements of the forward model A can be summarized as a contributions from (8.22, 8.23, 8.26, 8.27). We note that the composite moments are shown in the formulation of A and b . However, it should be noted that the composite moments do not occur in either of these variables. Instead,

I. Elements of matrix A : From (8.22)

(a) Elements stored in (row m , column m)

$$[\varphi_1]_m \left(\frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p + \frac{i\omega}{v} [\Delta V]_p + \sum_{i=1}^{S_I} \mathfrak{D}_1 \frac{dA_i}{dr_i} \right)$$

(b) Elements stored in (row m , column i)

$$[\varphi_1]_i \left(-\mathfrak{D}_1 \frac{dA_i}{dr_i} \right) \text{ (For each interior surface } i = 1 \dots S_I)$$

(c) Elements stored in (row m , column $m + M$)

$$[\varphi_2]_m \left(-\frac{2}{3} \frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p - \frac{2}{3} \frac{i\omega}{v} [\Delta V]_p \right)$$

II. Elements of matrix A : From (8.26)

(a) Elements stored in (row m , column m)

$$[\varphi_1]_m \left(\frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p + \frac{i\omega}{v} [\Delta V]_p + \sum_{i=1}^{S_B} \xi_{11} dA_i \right) \text{ (For each boundary surface } i = 1 \dots S_B)$$

(b) Elements stored in (row m , column $m + M$)

$$[\varphi_2]_m \left(-\frac{2}{3} \frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p - \frac{2}{3} \frac{i\omega}{v} [\Delta V]_p + \sum_{i=1}^{S_B} \xi_{12} dA_i \right) \text{ (For each boundary surface } i = 1 \dots S_B)$$

III. Elements of vector b : From (8.22, 8.26)

(a) Interior Eq. (row m)

$$[Q]_m [\Delta V]_p$$

(b) Boundary Eq. (row m)

$$[Q]_m [\Delta V]_p + \left(\xi_{13}[S_1]_p + \xi_{14}[S_2]_p \right) \sum_{i=1}^{S_B} dA_i \text{ (For each boundary surface } i = 1 \dots S_B)$$

Contributions from (8.23, 8.27).

I. Elements of matrix A : From (8.23)

(a) Elements stored in (row $m + M$, column m)

$$[\varphi_1]_m \left(-\frac{2}{3} \frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p \right)$$

(b) Elements stored in (row $m + M$, column $m + M$)

$$[\varphi_2]_m \left(\frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p + (1 - g^2) \frac{5}{9} \frac{[\hat{\mu}_s]_m - c_s}{l_s} [\Delta V]_p + \frac{1}{3} \frac{i\omega}{v} [\Delta V]_p + \sum_{i=1}^{S_I} \mathfrak{D}_2 \frac{dA_i}{dr_i} \right)$$

(c) Elements stored in (row $m + M$, column $i + M$)

$$[\varphi_2]_i \left(-\mathfrak{D}_2 \frac{dA_i}{dr_i} \right) \text{ (For each interior surface } i = 1 \dots S_I)$$

II. Elements of matrix A : From (8.27)

(a) Elements stored in (row $m + M$, column m)

$$[\varphi_1]_m \left(-\frac{2}{3} \frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p + \sum_{i=1}^{S_B} \xi_{21} dA_i \right) \text{ (For each boundary surface } i = 1 \dots S_B)$$

(b) Elements stored in (row $m + M$, column $m + M$)

$$[\varphi_2]_m \left(\frac{[\hat{\mu}_a]_m - c_a}{l_a} [\Delta V]_p + (1 - g^2) \frac{5}{9} \frac{[\hat{\mu}_s]_m - c_s}{l_s} [\Delta V]_p + \frac{1}{3} \frac{i\omega}{v} [\Delta V]_p + \sum_{i=1}^{S_B} \xi_{22} dA_i \right) \text{ (For each boundary surface } i = 1 \dots S_B)$$

III. Elements of vector b : From (8.23, 8.27)

(a) Interior Eq. (row $m + M$)

$$-\frac{2}{3}[Q]_m[\Delta V]_p$$

(b) Boundary Eq. (row $m + M$)

$$-\frac{2}{3}[Q]_m[\Delta V]_p + \left(\xi_{23}[S_1]_p + \xi_{24}[S_2]_p \right) \sum_{i=1}^M dA_i \text{ (For each boundary surface } i = 1 \dots S_B)$$

8.3.2 Partial current

The partial current, $J^+(\mathbf{r})$, is formally computed as

$$J^+(\mathbf{r}, \omega) = \mathbf{Q} [x^\top \ b^\top]^\top, \quad (8.33)$$

where \mathbf{Q} is a projection operator that transforms the composite moments φ_n and source terms $Q(\mathbf{r})$ and $S(\mathbf{r}, \Omega)$ into partial current measurements. For the SP_3 model, the projection operator \mathbf{Q} is given by

$$\mathbf{Q} = [Q_0 \ Q_1 \ Q_2 \ Q_3].$$

Matrix \mathbf{Q} can be derived from (8.28-8.29) and has dimensions $p \times 4M$, where p is the total number of source-detector pairs. Matrices $Q_n \in \mathbb{R}^{p \times M}$ have the following structure,

$$Q_n = \begin{bmatrix} \nu_n & 0 & \cdots & 0 & 0 \\ 0 & \nu_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \nu_n & 0 \\ 0 & 0 & \cdots & 0 & \nu_n \end{bmatrix}.$$

Each ν_n term is computed for each mesh node that maps to a detector.

8.3.3 Numerical solver for the forward model

The system of linear equations resulting from the FV-FD-SP_N model (8.30) is solved with a Krylov subspace iterative method famous for its efficiency in solving systems of linear equations known as the (restarted) generalized minimal residual method (GMRES) [203, 204]. GMRES is an attractive iterative solver as it can properly handle the forward model, where the coefficient matrix A is complex-valued and symmetric. The GMRES algorithm is reproduced in Algorithm 1 as it is presented in [204].

Algorithm 1 GMRES [204]

```

Compute  $r_0 = b - Ax_0$ ,  $\beta := \|r_0\|_2$ , and  $v_1 := r_0/\beta$ 
for  $j = 1, 2, \dots, m$  do
  Compute  $w_j := Av_j$ 
  for  $i = 1, \dots, j$  do
     $h_{ij} := (w_j, v_i)$ 
     $w_j := w_j - h_{ij}v_i$ 
  end for
   $h_{j+1,j} = \|w_j\|_2$ 
  if  $h_{j+1,j} = 0$  then
    Set  $m := j$  and go to last line
  end if
   $v_{j+1} = w_j/h_{j+1,j}$ 
end for
Define the  $(m+1) \times m$  Hessenberg matrix  $\bar{H}_m = \{h_{ij}\}_{1 \leq i \leq m+1, 1 \leq j \leq m}$ .
Compute  $y_m$  the minimizer of  $\|\beta e_1 - \bar{H}_m y\|_2$  and  $x_m = x_0 + V_m y_m$ .

```

The solution to the linear system (8.30) is given by x_m in the GMRES algorithm. It is worthwhile noting that the coefficient matrix A , and vectors x and y are complex-valued. As a result, the Hessenberg matrix (\bar{H}), the associated vectors (r, v, w), and the coefficient h are all complex-valued.

The derivation of the FVM algorithm for the alternate version of the FD-SP_N model [the FD-SP_N (CD) model in (8.3, 8.4, 8.7, 8.8, 8.10)], follows the same logic presented for FD-SP_N (RD). The only differences are the locations throughout the equations where the imaginary terms, $\frac{i\omega}{v}$, and its preceding weights occur.

8.4 Validation of FV-FD-SP_N algorithm

8.4.1 Numerical phantoms

The fluence (ϕ) and partial current (J^+) measurements obtained with the FV-FD-SP_N model are compared to solutions obtained with a benchmark algorithm based on the ERT model using a dense set of discrete ordinates (S_{12}). To differentiate between solutions from the SP_N (RD) (i.e. the model has strictly real-valued diffusion coefficients) and SP_N (CD) (i.e. the model has complex-valued diffusion coefficients) models, we use the following nomenclature: solutions computed with SP_N (RD) are labeled “SP_N (RD)” and solutions computed with SP_N (CD) are labeled “SP_N (CD).”

For numerical simulations we use a two-dimensional circular phantom with a diameter of 2.0 cm whose origin is at the center of the disk ($x = 0.0$ cm, $y = 0.0$ cm). The performances of the SP_N models are analyzed by varying the optical properties of the medium. Four specific sets of optical properties are considered, which together provide insight into the appropriateness of the various SP_N models in the extremely diffuse regime and in the non-diffuse regime.

The phantom is discretized by a dense numerical grid to ensure the numerical error from the ERT algorithm is minimal. The FEM mesh consists of 37,247 mesh nodes and 73,236 triangular

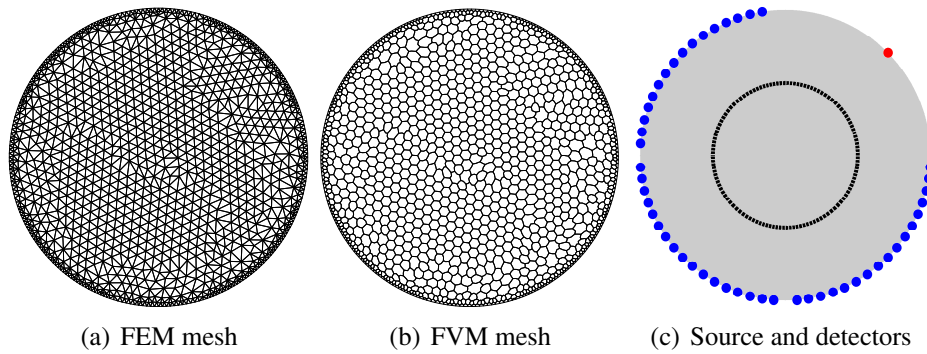


Figure 8.3: Discretization of a disk (2.0 cm diameter) into (a) triangular elements (i.e. FEM mesh) and (b) into finite-volume elements (i.e. FVM mesh). (c) A single boundary source (red dot) and 50 detectors (blue dots) are distributed around the boundary of the phantom. The internal cross-section of interest consists of mesh points 0.5 ± 0.01 cm from the center of the phantom, represented by the inner circle.

Table 8.4: Background properties of the numerical phantoms. The modulation frequency is from the relation $\omega = 2\pi f$.

Case	Optical properties characteristics	μ_a [cm ⁻¹]	μ_s [cm ⁻¹]	g	n_m	ω [MHz]
1	Homogeneous	0.001	400.0	0.95	1.0	600
2	Absorbing inclusions	0.1	100.0	0.95	1.0	600
3	Scattering inclusions	0.1	100.0	0.95	1.0	600
4	Absorbing and scattering inclusions	0.1	200.0	0.95	1.0	600

elements, resulting in an FVM mesh with 37,247 volume elements and 220,964 surfaces. The mesh is not shown as it is too dense to properly display, however, a significantly less dense mesh of the same phantom is shown [Fig. 8.3(a,b)]. The FEM mesh and FVM mesh are presented for comparison. In the example, the FEM mesh consists of 1,309 FEM nodes (2,360 triangles), while the FVM mesh contains 1,309 finite volume elements with 6,028 surfaces. The computation time required to generate the FVM mesh from the FEM mesh is minimal (0.011s in the example presented) and generally ignored.

There is a source on the boundary of the disk at position ($x = 0.7071$ cm, $y = 0.7071$ cm) with source power $S(\Omega) = 1.0$ W cm⁻² sr⁻¹ and 50 detectors distributed evenly around the phantom boundary at points away from the source (Fig. 8.3c). The partial current, $J^+(\mathbf{r})$, is computed at each of the detectors. The fluence, $\phi(\mathbf{r})$, is computed on all 37,247 FVM elements. To analyze the differences between the fluence computed with the SP_N models and the benchmark solution, a cross-section of interest is defined within the phantom, which includes all mesh points 0.5 ± 0.01 cm from the center of the phantom (Fig. 8.3c).⁵

The background optical properties of the four phantoms are summarized in Table 8.4 and the specific location of the inclusions and their optical properties are shown in Fig. 8.4. The first phantom (Fig. 8.4a) is homogeneous with low absorption ($\mu_a = 0.001$ cm⁻¹) and high scattering ($\mu_s = 400$ cm⁻¹, $g = 0.95$). This medium simulates highly diffuse tissue where the diffusion equation is known to be an accurate model of light propagation.

⁵This cross-section of interest is chosen as it crosses all inclusions and, therefore, allows for the analysis of the fluence within those objects of interest.

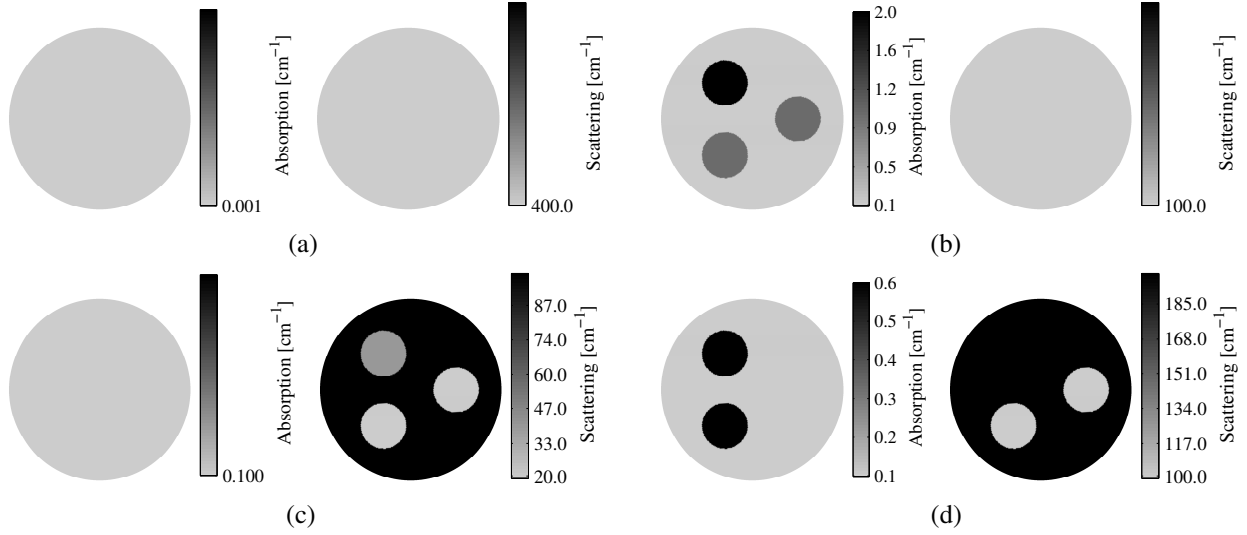


Figure 8.4: (a-d) Definition of two-dimensional phantoms number 1 through 4, respectively. (a) Homogeneous and (b-c) inhomogeneous phantoms. All simulations on these phantoms are performed with anisotropic factor $g = 0.95$, source modulation frequency $\omega = 2\pi f = 600$ MHz, and refractive index $n_m = 1.0$.

The remaining three phantoms are inhomogeneous distributions of optical properties with either highly absorbing inclusions (phantoms number two, Fig. 8.4b), low scattering inclusions (phantoms number three, Fig. 8.4c), or a combination of these inclusions (phantoms number four, Fig. 8.4d). Each of the inclusions are disks with 0.25 cm diameter. Phantoms number two, three, and four are designed to test the accuracy of the SP_N equations in the transport regime, where the diffusion equation is known to be a poor approximation to the ERT.

The relative error, $J_e^+(\mathbf{r})$, between the benchmark partial current, $J_{ERT}^+(\mathbf{r})$, and the partial current computed with the SP_1 and SP_3 models, $J_{SP1}^+(\mathbf{r})$ and $J_{SP3}^+(\mathbf{r})$, is computed at each mesh node i using the following formula:

$$J_e^+(\mathbf{r})_i = 100 \times \frac{J_{SPN}^+(\mathbf{r})_i - J_{ERT}^+(\mathbf{r})_i}{J_{ERT}^+(\mathbf{r})_i} \quad (8.34)$$

The *average* relative error of the partial current, $\tilde{J}_e^+(\mathbf{r})$, is also reported as a measure of overall error. The relative error, $\phi_e(\mathbf{r})$, between the benchmark fluence, $\phi^{ERT}(\mathbf{r})$, and the 0th order

moment of the radiance (i.e. the fluence) computed with the SP_1 and SP_3 models, $\phi^{SP1}(\mathbf{r})$ and $\phi^{SP3}(\mathbf{r})$, is computed at each mesh node i using (8.35). A second parameter that quantifies the error between the SP_N and ERT solutions is the average error between the solutions at each mesh point, denoted $\tilde{\phi}_e(\mathbf{r})$, and is given by (8.36). The finite volume ΔV (or finite area in the case of two-dimensional problems) of each FVM element is used as a weight factor in the Riemann sum.

$$\phi_e(\mathbf{r})_i = 100 \times \frac{\phi^{SPN}(\mathbf{r})_i - \phi_i^{ERT}(\mathbf{r})_i}{\phi_i^{ERT}(\mathbf{r})_i} \quad (8.35)$$

$$\tilde{\phi}_e(\mathbf{r}) = \frac{100}{V} \times \sum_{i=1}^N \left(\frac{\phi^{SPN}(\mathbf{r})_i - \phi_i^{ERT}(\mathbf{r})_i}{\phi_i^{ERT}(\mathbf{r})_i} \right) \Delta V_i \quad (8.36)$$

Here, V represents the total computational volume (or total area in two-dimensional problems). Error measures equivalent to (8.35-8.36) between the fluence computed with the SP_N and ERT along the internal cross-section (Fig. 8.3c) are denoted as $\phi_e^\bullet(\mathbf{r})_i$ and $\tilde{\phi}_e^\bullet(\mathbf{r})$, respectively.

The restarted GMRES algorithm is used to solve the forward problem in all cases. In these examples, the GMRES algorithm is set to restart after 30 iterations with no limit on the number of total iterations. The convergence criteria is set to 1×10^{-12} .

With the SP_3 model, the photon fluence (ϕ_0) throughout the medium is obtained by a linear combination of the first two composite moments, φ_1 and φ_2 ($\phi_0 = \varphi_1 - \frac{2}{3}\varphi_2$). The individual moments are presented for all simulations with the SP_3 model for detailed analysis.

8.4.2 Numerical study - phantom 1: homogeneous media

Results from simulations on the homogeneous phantom with highly diffuse properties (Fig. 8.4a) are presented in this section. All three light propagation models return fluence distributions that are qualitatively similar, verifying that all models converge to similar solutions in highly diffuse media. Images of the amplitude and the phase of the fluence computed with the SP_N models and the ERT model are presented in Figs. 8.5-8.6. Maps of the relative error between the SP_N solutions and the benchmark ERT solution are also presented in these figures. The average errors of the fluence and the partial current are summarized in Table 8.5.

The relative error of the amplitude of the fluence, Fig. 8.5(f-i), shows agreement between all models. The average relative error between the SP_N solutions and the benchmark solution is between 9.90% and 10.51%. At the individual mesh node level, the error is typically in the range of $\pm 20.0\%$, however, the error at mesh nodes near the source can reach much higher values and has been artificially capped at $\pm 20.0\%$ in order to properly display the error across the rest of the phantom. Thus, all SP_N models are incapable of accurately modeling the “transport” behavior of photons near the source. However, because the media is highly diffuse, transport photons become diffuse after only a short distance, which is the reason why the error in the SP_N solutions is lower (generally bounded by $\pm 20.0\%$) across the rest of the medium.

Similar observations can be made from analysis of the phase of the fluence (Fig. 8.6), again suggesting there is significant agreement between all solutions. The average relative error between the phase of the fluence ranges from 9.32% to 10.04%. While there is significant error at mesh

Table 8.5: Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 1.

Model	$\tilde{\phi}_e(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e(\mathbf{r})$ [%] (Phase)	$\tilde{\phi}_e^*(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e^*(\mathbf{r})$ [%] (Phase)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Amplitude)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Phase)
SP_1 (CD)	9.90	9.65	8.89	9.76	3.85	10.72
SP_1 (RD)	10.51	9.32	9.55	9.47	4.29	10.09
SP_3 (CD)	9.96	10.04	9.30	10.09	4.10	10.85
SP_3 (RD)	10.31	9.64	9.68	9.77	4.35	10.36

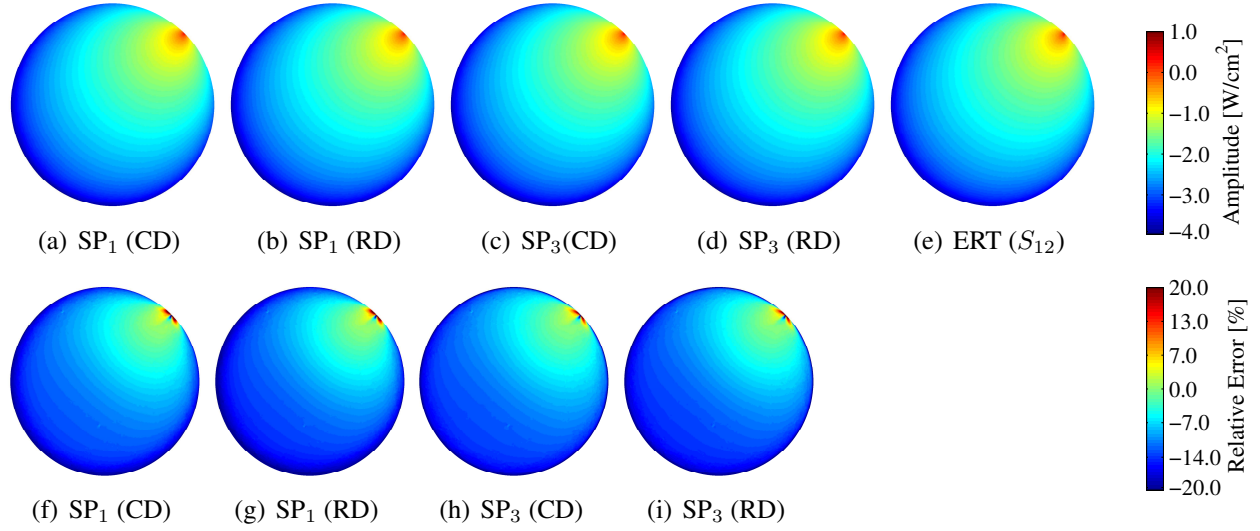


Figure 8.5: [Phantom 1] Amplitude (log scale) of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

nodes within close proximity to the source, the error at nodes a short distance from the node is significantly lower.

The individual components of the fluence computed with the SP_3 model (i.e. the first two composite moments φ_1 and φ_2) are presented in Fig. 8.7. The amplitude and the phase distributions of φ_1 are similar for both CD and RD methods, although the SP_3 (CD) model exhibits marginally lower overall relative error in amplitude (9.96%) than the SP_3 (RD) model (10.31%). However, there are significant differences between the amplitude and phase distributions of φ_2 . The CD model produces a φ_2 with significantly larger amplitude, suggesting that the CD model “corrects” φ_1 more than the RD model. It is interesting to note that this does not actually result in improved solutions. Instead, the SP_3 (CD) model exhibits higher overall relative error in phase (10.04%) than the SP_3 (RD) model (9.64%).

The partial current at the boundary (Fig. 8.8) is approximated to similar accuracies by all SP_N models. The average error of the amplitude of the partial current ranges from 3.85% to 4.35%. The average error between the phase of the partial current is larger, ranging from 10.09% to 10.85%.

The fluence within the internal cross-section of interest (Fig. 8.9) also supports the observation that all SP_N models exhibit similar performance in the diffuse phantom. The errors in the amplitude

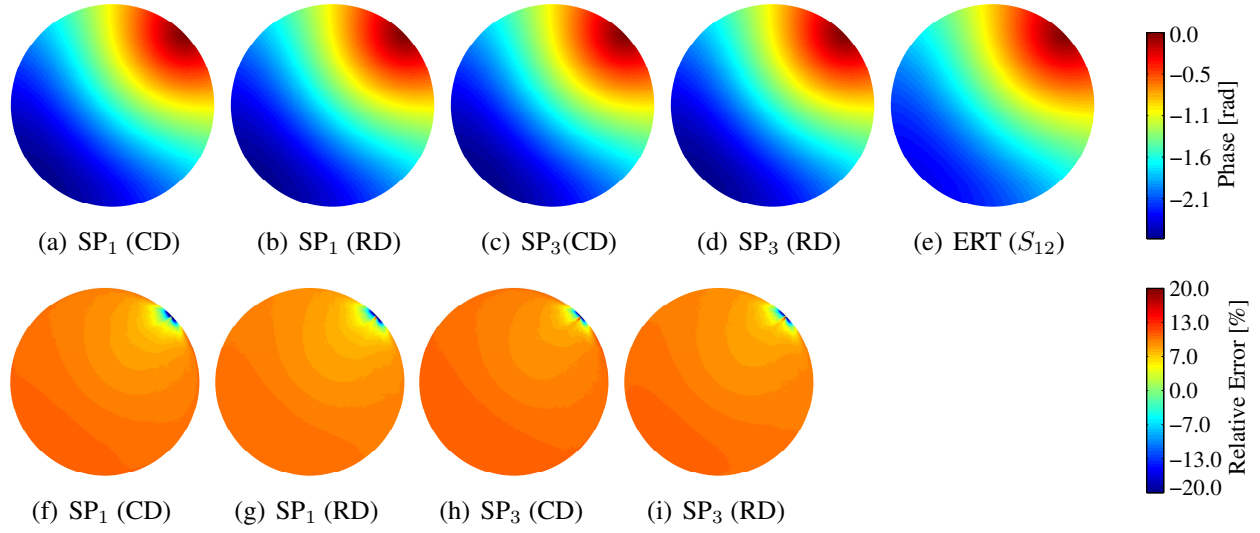


Figure 8.6: [Phantom 1] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

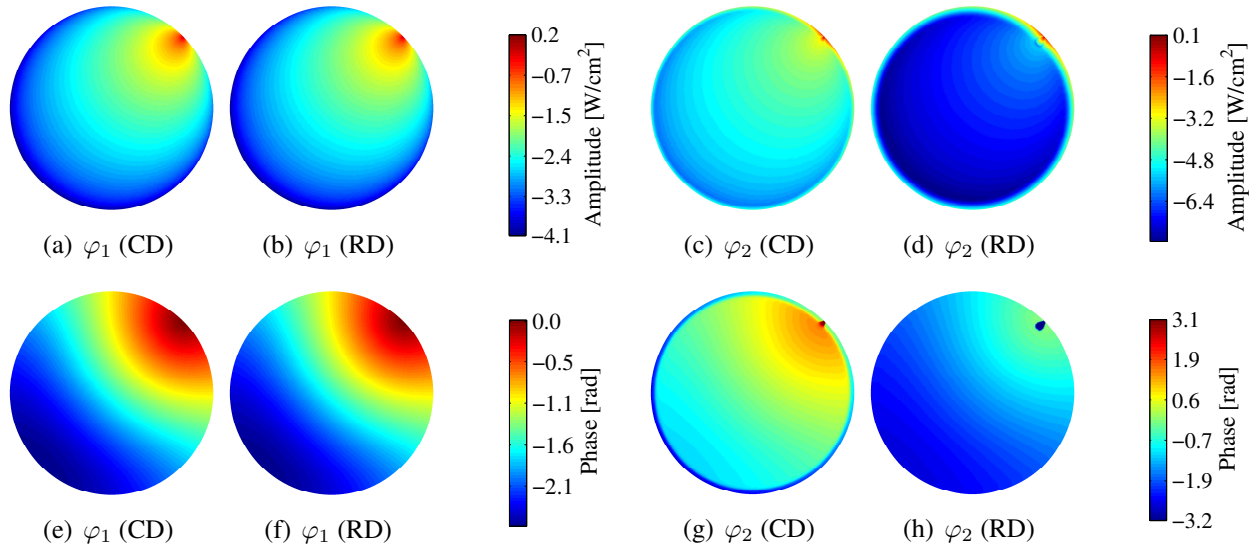


Figure 8.7: [Phantom 1] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2 .

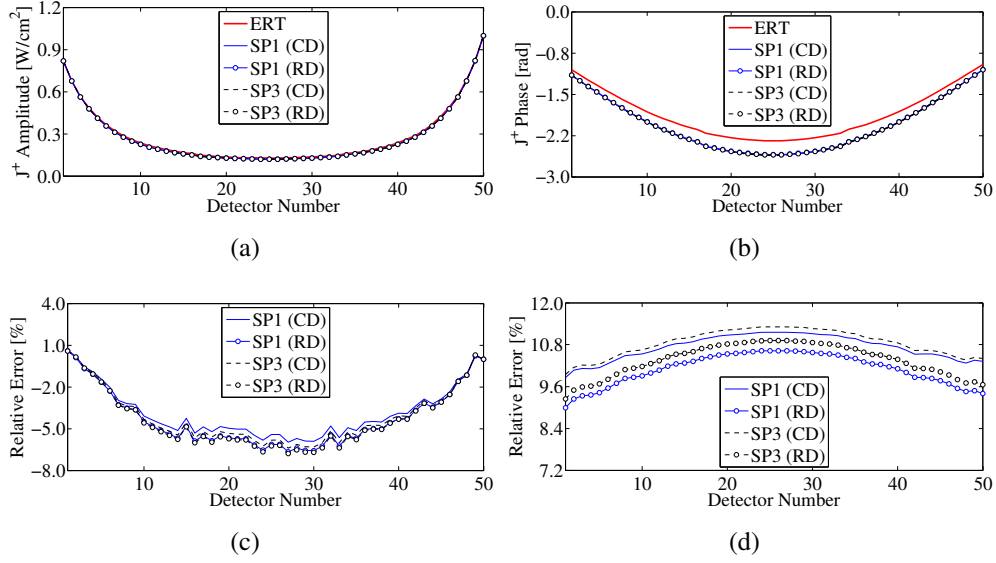


Figure 8.8: [Phantom 1] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.

Table 8.6: Computation time for phantom 1.

Model	Time [s]	RAM [MB]
SP_1 (RD)	53.5	172.6
SP_1 (CD)	76.6	194.3
SP_3 (RD)	115.9	313.1
SP_3 (CD)	150.3	359.4
ERT (S_{12})	1188.0	6,140.0

of the fluence from all SP_N models are similar, with the average error $\tilde{\phi}_e^\bullet(\mathbf{r})$ ranging from 8.89% to 9.68%, while the average error in the phase of the fluence ranges from 9.47% to 10.09%.

The computation time required to obtain the fluence with each of the models make the benefits of the SP_N model over the ERT model clear (Table 8.6). Relative to the computational time required to obtain the fluence with the SP_1 model, the SP_3 takes 2.06 times longer, while the ERT (S_{12}) solution requires 184.5 times more computation time. These results are consistent with our expectations, as the SP_3 model results in a system of linear equations twice as large as the SP_1 system of equations.

Analysis of the performance of the SP_N models in a diffuse media reveals that they all perform

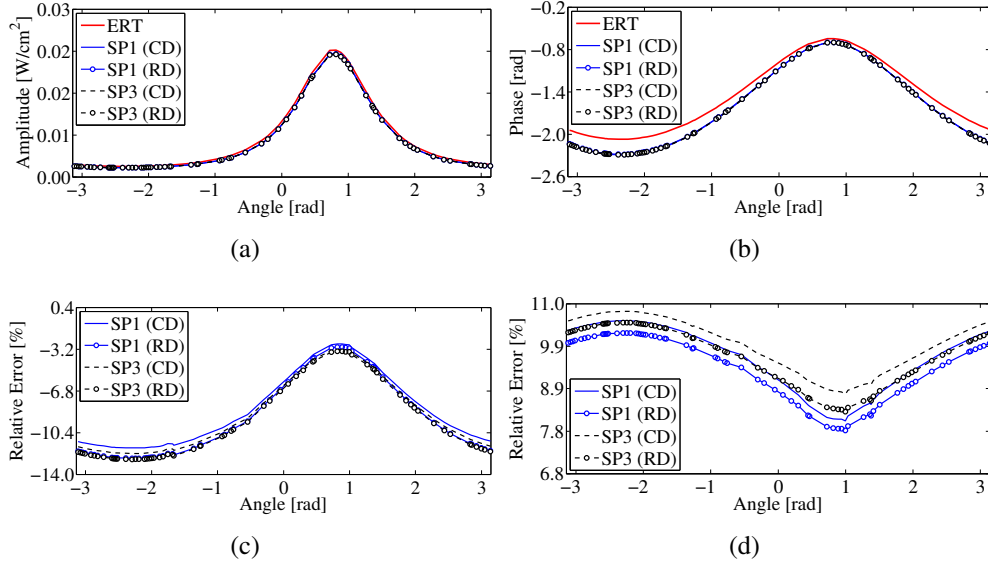


Figure 8.9: [Phantom 1] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.

similarly well in approximating the ERT.

8.4.3 Numerical study - phantom 2: absorbing inclusions

Results from simulations on the phantom that includes highly absorbing inclusions (Fig. 8.4b) are presented in this section. There are three objects in the medium; two objects with $\mu_a = 1.0 \text{ cm}^{-1}$ and one object with $\mu_a = 2.0 \text{ cm}^{-1}$. The average relative errors are summarized in Table 8.7. Images of the amplitude and the phase of the fluence are presented in Figs. 8.10 and 8.11, respectively. The individual composite moments of the SP_3 models are shown in Fig. 8.12. The fluence on the cross-section of interest and the partial current at the boundary detectors are presented in Figs. 8.13 and 8.14, respectively.

The fluence and the relative error of the fluence (Fig. 8.10) show that the SP_3 models are significantly better than the SP_1 models at approximating the benchmark solution. In general, all SP_N solutions exhibit large positive errors near the source (overestimation of the fluence) and negative errors far from the source (underestimation of the fluence). While both models exhibit large errors near the source, the area within which large errors occur is larger in the SP_1 solutions than in the SP_3 solutions. Similarly, the SP_3 solutions are more accurate approximations to the benchmark than the SP_1 solutions at locations far from the boundary.

Furthermore, it is clear that the SP_3 models are able to more accurately approximate the benchmark solution in the region of the highly absorbing inclusions, as indicated by the presence of pronounced areas of error in the SP_1 solutions at the location of the inclusions. The error profiles of the SP_3 solutions in these same regions are less pronounced, indicating that the SP_3 equations model light propagation in highly absorbing media more accurately than the SP_1 equations.

Table 8.7: Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 2.

Model	$\tilde{\phi}_e(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e(\mathbf{r})$ [%] (Phase)	$\tilde{\phi}_e^\bullet(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e^\bullet(\mathbf{r})$ [%] (Phase)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Amplitude)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Phase)
SP_1 (CD)	8.14	5.44	4.09	3.15	15.21	8.68
SP_1 (RD)	8.51	2.93	4.62	1.77	15.53	1.61
SP_3 (CD)	5.26	4.27	3.89	3.46	11.64	6.27
SP_3 (RD)	5.47	2.54	4.15	1.43	11.82	1.51

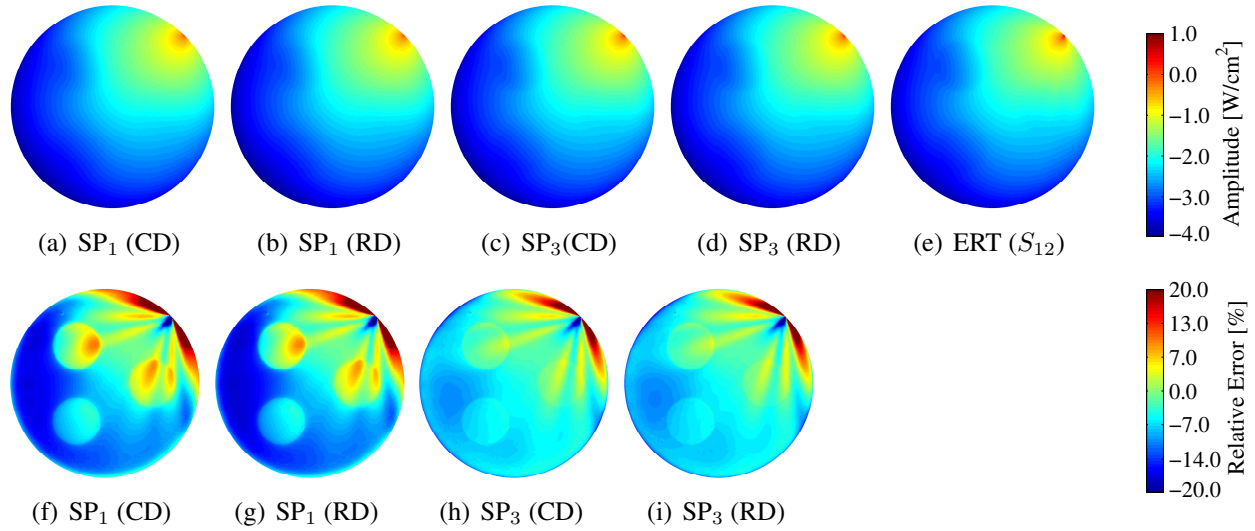


Figure 8.10: [Phantom 2] Amplitude of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

Both SP_3 models (CD and RD) produce more accurate solutions than the two SP_1 models (CD and RD). Additionally, both SP_3 models (CD and RD) exhibit similar overall relative error in the amplitude (5.26% and 5.47%, respectively), but differ significantly in the error of the phase.

Analysis of the phase of the fluence reinforces the observation that the SP_3 models are better approximations to the ERT solution than the SP_1 models (Fig. 8.11). All models produce phase distributions that significantly underestimate the phase near the source and moderately overestimate it far from the source. However, the area around the source within which the SP_3 models underestimate the source is significantly smaller than the area in which the SP_1 models severely understate the phase. This suggests that the “transport” regime, within which the phase accumulation of photon propagation can only be modeled by the ERT, is smaller for the SP_3 models than the SP_1 models. We note that the pronounced areas of error that appear in the amplitude of the fluence at the location of the highly absorbing inclusions are not present in the relative error of the phase.

The average relative error of the phase of the SP_3 (RD) solution (2.54%) is significantly smaller than the average relative error of the phase of the SP_3 (CD) solution (4.27%). This observation is in contrast to the pattern observed in the amplitude of the phase, where the differences between the SP_3 (RD) and SP_3 (CD) solutions are minimal (5.47% and 5.26%, respectively). Surprisingly,

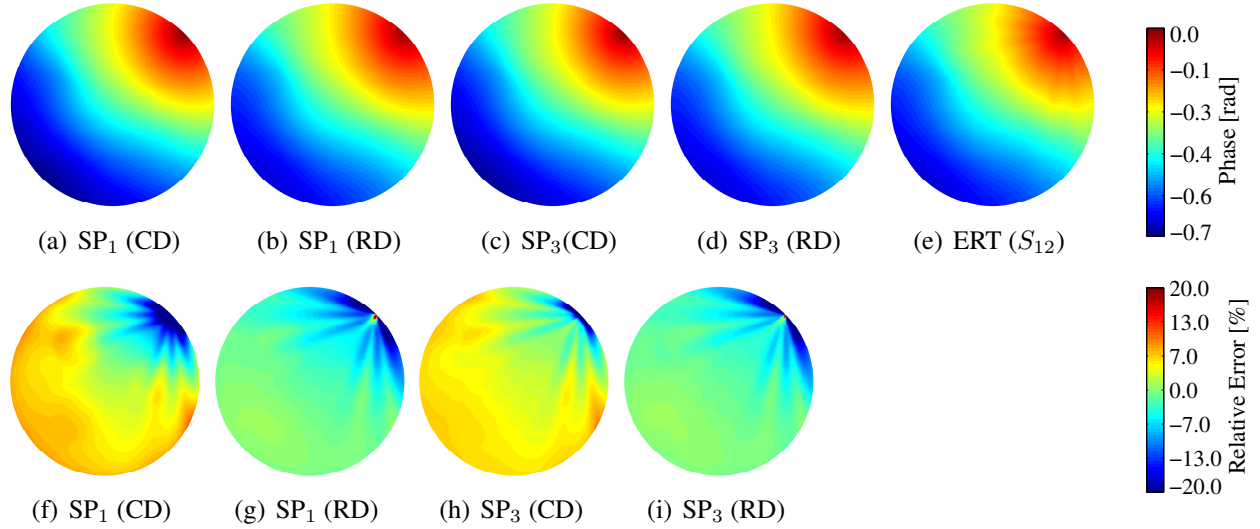


Figure 8.11: [Phantom 2] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

the SP_1 (RD) model achieves a lower average relative error of the phase (2.93%) than the SP_3 (CD) model (4.27%). This suggests that the SP_1 (RD) model is more accurate than the SP_1 (CD) and SP_3 (CD) models in modeling phase accumulation. The SP_N (RD) solutions, however, have a marginally bigger area of large error near the boundary when compared to the SP_N (CD) solutions.

Analysis of the composite moments of the radiance produced by the SP_3 models is informative (Fig. 8.12). The amplitude and the phase of the first composite moment (φ_1) are similar for both CD and RD models [Fig. 8.12(a,b,e,f)]. There are small perturbations of the solution in the region of the inclusions. The second moment (φ_2) is particularly interesting because the amplitude of φ_2 is prominently higher in the exact locations of the highly absorbing inclusions, making it exceptionally clear where the inclusions are located [Fig. 8.12(c,d)]. This is interesting as it may be possible to exploit this information to improve the image reconstruction process. For example, this knowledge may be used to dynamically increase mesh density in the areas with prominent “correction” terms (i.e. as observed in φ_2).

It is also interesting to note that there are significant differences between the CD and RD models in the phase of φ_2 , as the location of the absorbing inclusions are clearly visible in φ_2 (CD) but not in φ_2 (RD) [Fig. 8.12(g,h)]. The value of φ_2 (CD) near the source is much larger than the values

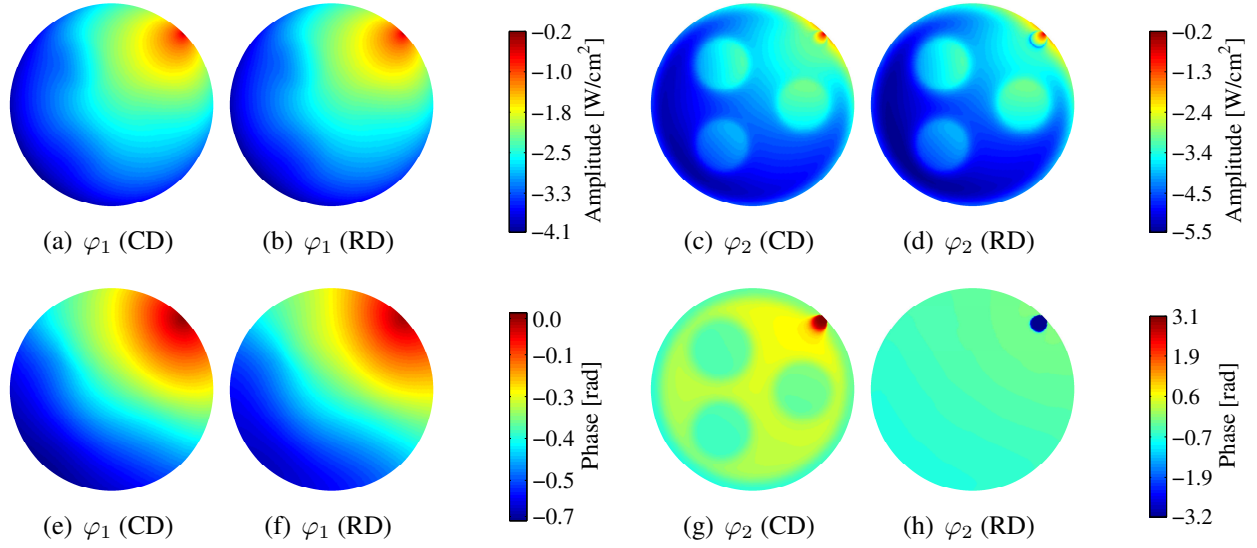


Figure 8.12: [Phantom 2] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2 .

at the same locations of φ_2 (RD). It is likely that this strong “correction” near the boundary is responsible for improved accuracy in the phase of the fluence computed with the SP_3 (CD) model at mesh nodes near the source. However, the strong “correction” in the interior of the φ_2 (CD) solution may also be the reason why mesh nodes far from the source see an increase in error in the phase of the fluence computed with the SP_3 (CD) model.

Thus, while φ_2 (CD) correction terms result in a reduction of the error in the phase of the fluence near the boundary, they are also responsible for the increase in error in the phase of the fluence far from the boundary.

Comparing the fluence along the internal cross-section (Fig. 8.13) supports the observations thus far. Both SP_3 (CD and RD) models are better approximations to the amplitude of the benchmark solution than the SP_1 models, with insignificant differences between the RD and CD models. In the case of the phase of the fluence, however, the SP_1 (RD) and SP_3 (RD) models are a better approximation than either the SP_1 (CD) or SP_3 (CD) models. This is in agreement with our observations in the analysis of the fluence not the entire phantom.

The partial current at the boundary (Fig. 8.14) supports these same observations. The SP_3 mod-

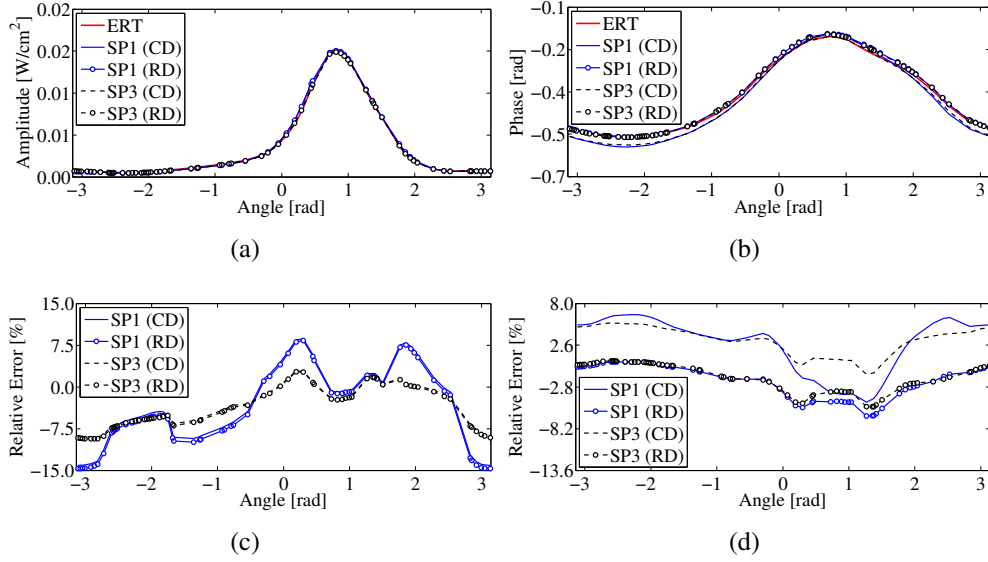


Figure 8.13: [Phantom 2] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.

els (CD and RD) predict the amplitude of $J^+(\mathbf{r})$ more accurately than the SP_1 models. However, the SP_1 (RD) and SP_3 (RD) models are better at approximating the phase of $J^+(\mathbf{r})$ than the SP_1 (CD) and SP_3 (CD) models.

The computation time and memory requirements necessary to solve for the fluence with each of the models makes the benefits of the SP_N models over the ERT model clear (Table 8.8). The computation time required to obtain the fluence using the SP_N (RD) models is significantly lower than the time required by the SP_N (CD) models (approximately 40% lower). The increase in computation time (and memory requirements) for SP_N (CD) is expected as all entries of the forward model operator A are complex-valued, while only the main diagonal and some off-diagonal elements are complex-valued in the RD system.

Overall, the SP_3 (RD) model performs better than all other SP_N models. Furthermore, the SP_1 (RD) model is more accurate than the SP_1 (CD) and SP_3 (CD) model on approximating the phase of the benchmark fluence and partial current values. This is not surprising, given that we expect SP_1 (RD) SP_N (RD) equations to be the more accurate model of light propagation.

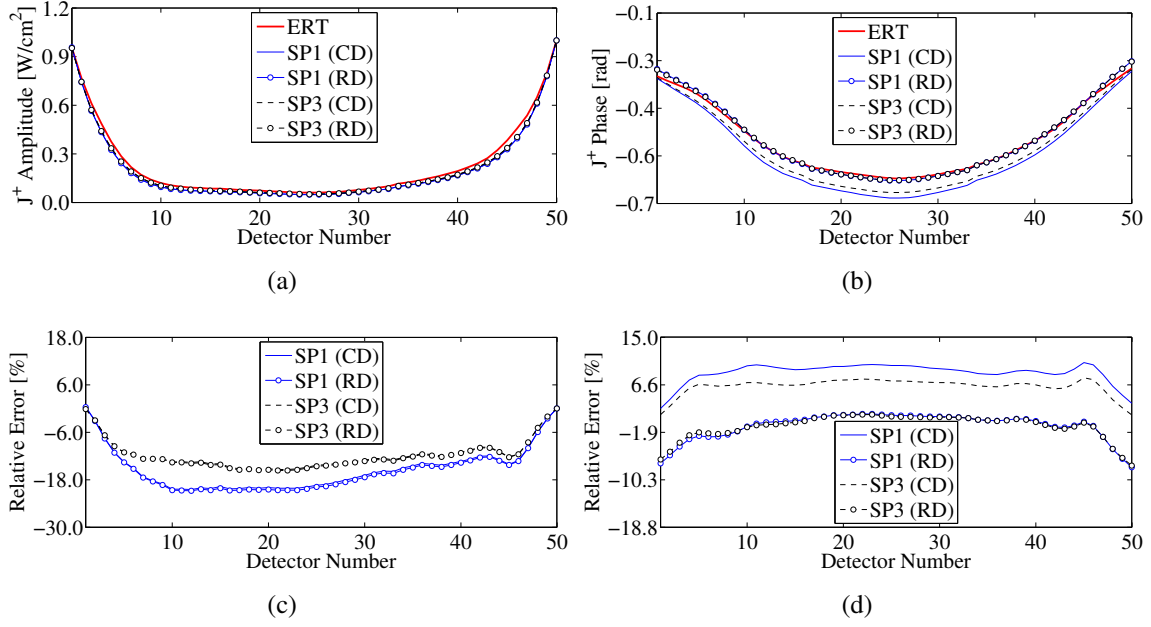


Figure 8.14: [Phantom 2] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.

Table 8.8: Computation time for phantom 2.

Model	Time [s]	RAM [MB]
SP_1 (RD)	30.9	172.6
SP_1 (CD)	49.1	194.3
SP_3 (RD)	63.2	313.1
SP_3 (CD)	100.1	359.4
ERT (S_{12})	4819.3	6,140.0

8.4.4 Numerical study - phantom 3: scattering inclusions

Results from simulations on the phantom with areas with low scattering values (Fig. 8.4c) are presented in this section. There are three objects in the medium; two objects with $\mu_s = 20.0 \text{ cm}^{-1}$ and one object with $\mu_s = 60.0 \text{ cm}^{-1}$. The average relative errors are summarized in Table 8.9. Images of the amplitude and the phase of the fluence are presented in Figs. 8.15 and 8.16, respectively. The individual composite moments of the SP_3 models are shown in Fig. 8.17. Plots of the fluence on the cross-section of interest and the partial current at the boundary detectors are presented in Figs. 8.18 and 8.19, respectively.

Overall, the SP_3 models perform better than the SP_1 models (Table 8.9). The error in the amplitude of the fluence obtained with both SP_3 models is similar (4.87% and 4.69%, respectively). The SP_3 (CD) models, however, yield solutions with lower relative error in the phase of the fluence (2.67%) compared to the SP_3 (RD) model (3.94%).

All SP_N solutions have large errors in the amplitude and the phase of the fluence near the source. The amplitude of the fluence is generally overestimated near the source (Fig. 8.15), while the phase is underestimated at these same locations (Fig. 8.16). Compared to the SP_3 solutions, the amplitude of the SP_1 solutions exhibit large positive errors across a larger area near the source. In addition, the amplitude of the SP_1 solutions are severely overestimated at the boundary directly opposite the source (Fig. 8.15).

The amplitude and phase of the first moment φ_1 is similar between the SP_3 (CD) and SP_3 (RD) models. However, there are significant differences between the phase profiles of φ_2 . The fluence

Table 8.9: Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 3.

Model	$\tilde{\phi}_e(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e(\mathbf{r})$ [%] (Phase)	$\tilde{\phi}_e^\bullet(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e^\bullet(\mathbf{r})$ [%] (Phase)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Amplitude)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Phase)
SP_1 (CD)	5.45	3.56	2.07	1.65	13.27	5.00
SP_1 (RD)	4.99	4.28	1.10	3.54	13.79	3.09
SP_3 (CD)	4.87	2.67	0.67	1.09	12.56	3.74
SP_3 (RD)	4.69	3.94	0.27	3.20	12.88	3.04

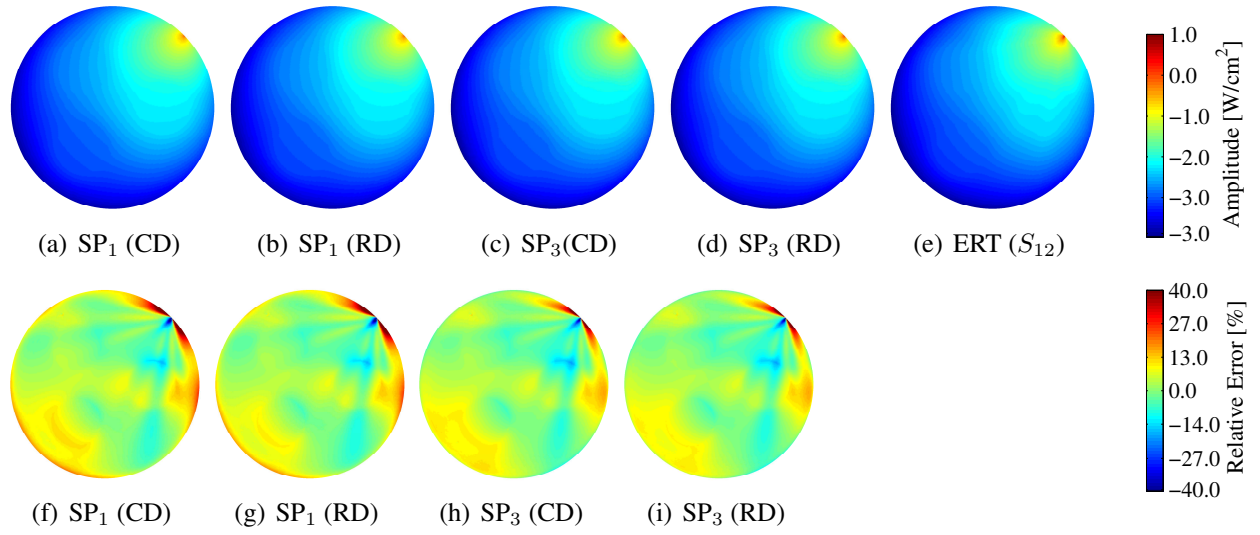


Figure 8.15: [Phantom 3] Amplitude of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

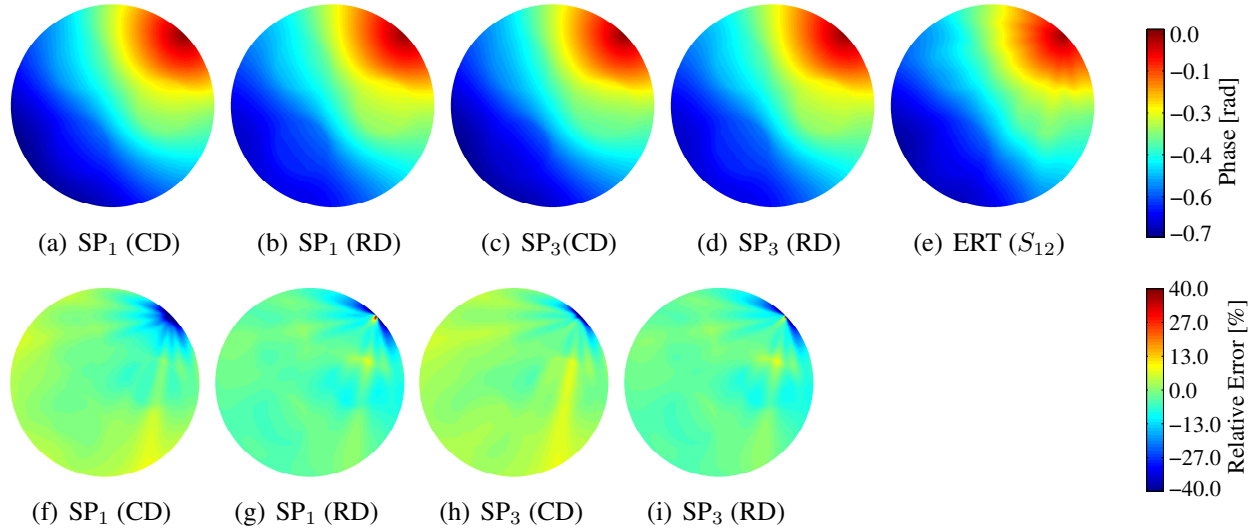


Figure 8.16: [Phantom 3] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

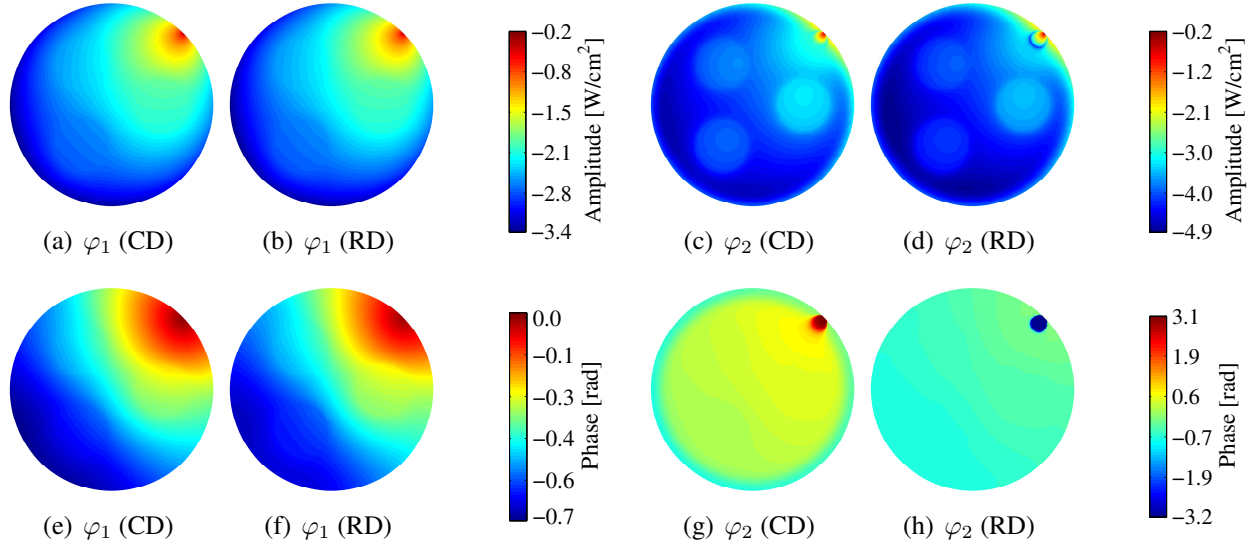


Figure 8.17: [Phantom 3] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP₃ (CD) and SP₃ (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2 .

along the internal cross-section (Fig. 8.18) of interest and the partial current (Fig. 8.19) support the same observations. The scattering inclusions are identifiable in the images of the amplitude of φ_2 [Fig.8.17(c,d)]. However, the scattering inclusions are not identifiable in the images of the phase of φ_2 [Fig.8.17(g,h)]. The contrast at the site of the scattering inclusions is lower than the contrast at the site of the absorbing inclusions considered in the second phantom, particularly in the amplitude images [Fig. 8.12(c,d)].

In general, the SP₃ models outperform the SP₁ model. The SP₃ (RD) model is better than the SP₃ (CD) model at approximating the phase of the fluence and partial current. However, the SP₃ (CD) model is better at approximating the amplitude of the benchmark solutions. The fluence along the internal cross-section (Fig. 8.18) of interest and the partial current (Fig. 8.19) support the same observations.

The computation times required to obtain the fluence using the SP_N (RD) models are approximately 70% lower than the time required by the SP_N (CD) models (Table 8.10). The increase in computation time (and memory requirements) associated with solving the SP_N (CD) equations is larger than what is observed when only highly absorbing inclusions are considered.

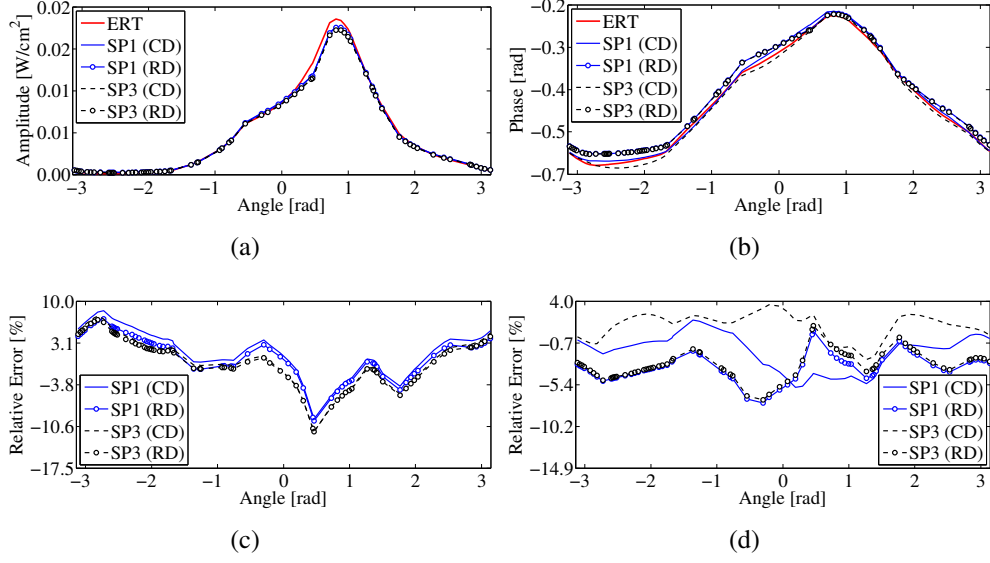


Figure 8.18: [Phantom 3] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.

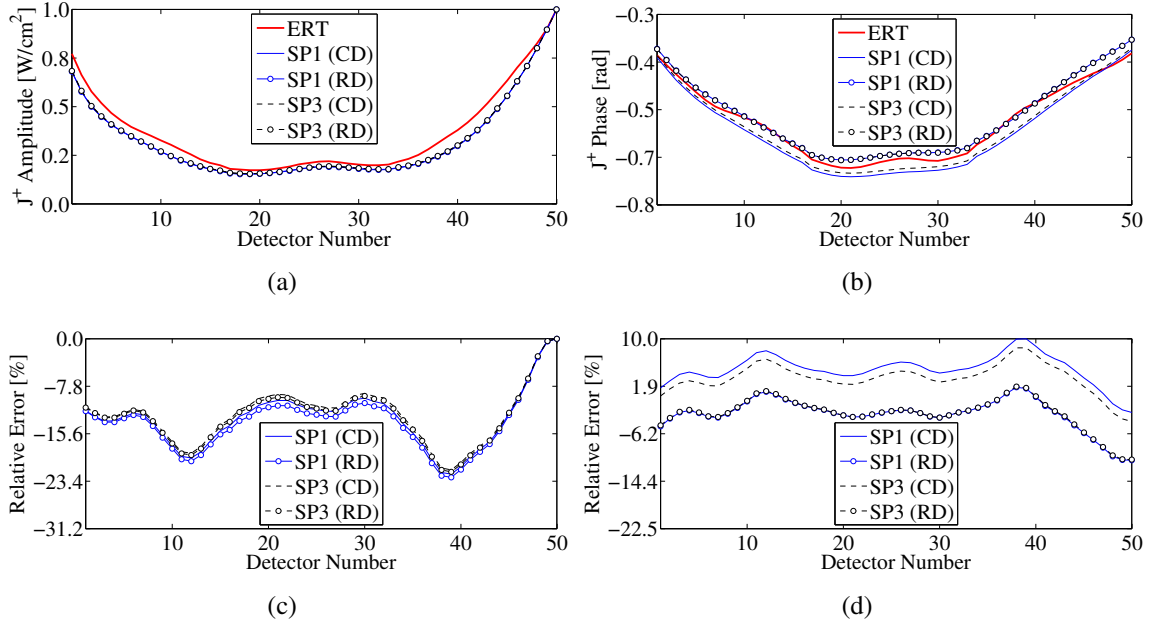


Figure 8.19: [Phantom 3] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.

Table 8.10: Computation time for phantom 3.

Model	Time [s]	RAM [MB]
SP ₁ (RD)	99.6	176.6
SP ₁ (CD)	333.2	194.3
SP ₃ (RD)	205.8	313.1
SP ₃ (CD)	684.6	359.4
ERT (S_{12})	7816.4	6,140.0

8.4.5 Numerical study - phantom 4: absorbing and scattering inclusions

Results from simulations on the fourth phantom, which contains absorbing and void like inclusions (Fig. 8.4d) are presented in this section. There are four objects in the medium; two objects with $\mu_a = 0.6 \text{ cm}^{-1}$ and two objects with $\mu_s = 100.0 \text{ cm}^{-1}$. The average relative errors are summarized in Table 8.11. Images of the amplitude and the phase of the fluence are presented in Figs. 8.20 and 8.21, respectively. The individual composite moments of the SP₃ models are shown in Fig. 8.22. Plots of the fluence on the cross-section of interest and the partial current at the boundary detectors are presented in Figs. 8.23 and 8.24, respectively.

All SP_N models approximate the amplitude of the ERT solution with similar accuracy. The SP_N (CD) models are marginally more accurate than the SP_N (RD) models (Fig. 8.20). However, the SP_N (RD) models perform significantly better than the SP_N (CD) models in approximating the phase of the benchmark solution (Fig. 8.21). The same trends are evident in the amplitude and the phase of the fluence along the internal cross-section of interest (Fig. 8.23).

Analysis of the partial current reveals similar trends (Fig. 8.24). The SP₃ models approximate

Table 8.11: Average relative error of the fluence, $\tilde{\phi}_e(\mathbf{r})$, and partial current, $\tilde{J}_e^+(\mathbf{r})$, for phantom 4.

Model	$\tilde{\phi}_e(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e(\mathbf{r})$ [%] (Phase)	$\tilde{\phi}_e^\bullet(\mathbf{r})$ [%] (Amplitude)	$\tilde{\phi}_e^\bullet(\mathbf{r})$ [%] (Phase)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Amplitude)	$\tilde{J}_e^+(\mathbf{r})$ [%] (Phase)
SP ₁ (CD)	5.84	4.89	5.10	4.60	7.88	7.06
SP ₁ (RD)	6.38	3.13	5.18	3.01	8.21	3.76
SP ₃ (CD)	5.72	5.02	5.53	4.95	6.80	6.29
SP ₃ (RD)	6.01	3.28	5.83	3.22	7.01	3.67

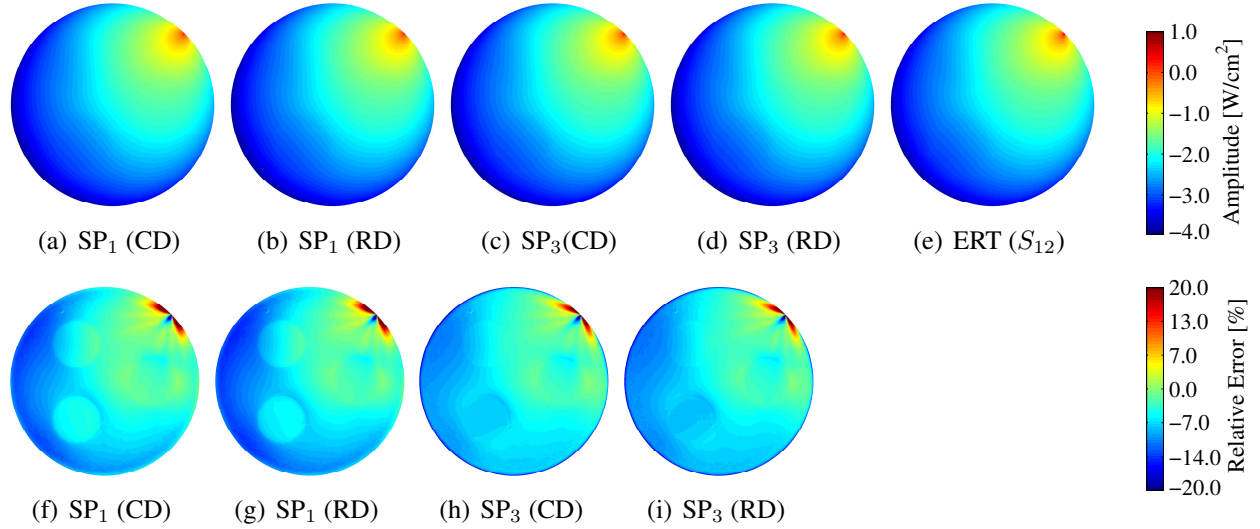


Figure 8.20: [Phantom 4] Amplitude of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

the benchmark partial current more accurately than their corresponding SP_1 models. However, the SP_3 (CD) model yields the most accurate amplitude of the partial current, while the SP_3 (RD) model best approximates the phase of the partial current.

The location of the absorbing objects are clearly visible in the phase of φ_2 (Fig. 8.22). Thus, while scattering and absorbing inclusions are undifferentiated in the amplitude of φ_2 , the phase of φ_2 clearly differentiates between the absorption and the scattering coefficients values in the inclusions. The amplitude and the phase of the first moment (φ_1) are similar for both CD and RD models, with the appearance of small perturbations of the solution in the region of the inclusions.

The SP_N (RD) models require less computation time compared to the SP_N (CD) models (Table 8.12). Furthermore, the SP_N (RD) models require less RAM than the SP_N (CD) models during the execution of the algorithm. The benefit of using any of the SP_N models is clear when considering the computation time (~ 150 times more than SP_1) and the memory requirements associated with using the ERT S_{12} model (≥ 6.14 GB).

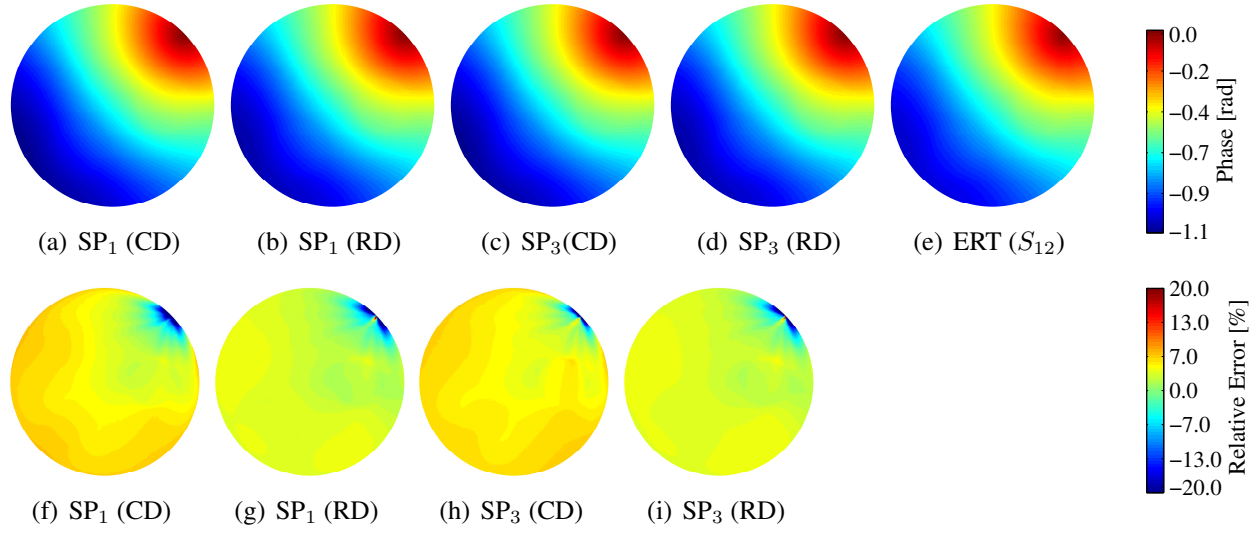


Figure 8.21: [Phantom 4] Phase of the fluence computed with the (a) SP_1 (CD), (b) SP_1 (RD), (c) SP_3 (CD), (d) SP_3 (RD), and (e) ERT (S_{12}) models. (Bottom row) Percent error at each node relative to the ERT benchmark (S_{12}).

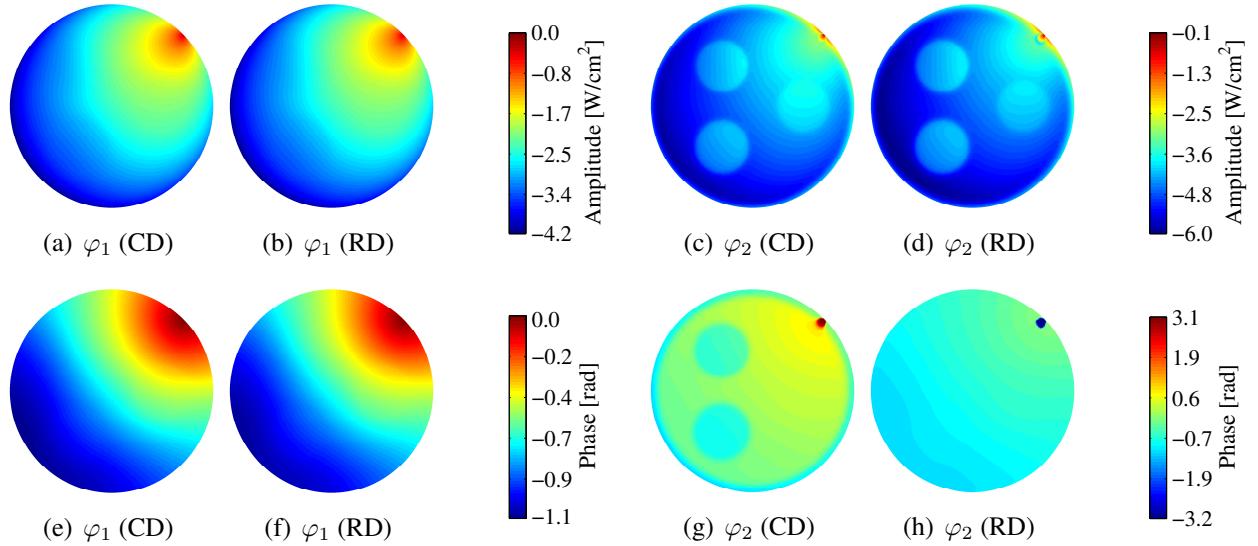


Figure 8.22: [Phantom 4] Amplitude and phase of the composite moments (φ_1 and φ_2) of the fluence computed with the SP_3 (CD) and SP_3 (RD) models. (a,b) Amplitude and (e,f) phase of φ_1 . (c,d) Amplitude and (g,h) phase of φ_2 .

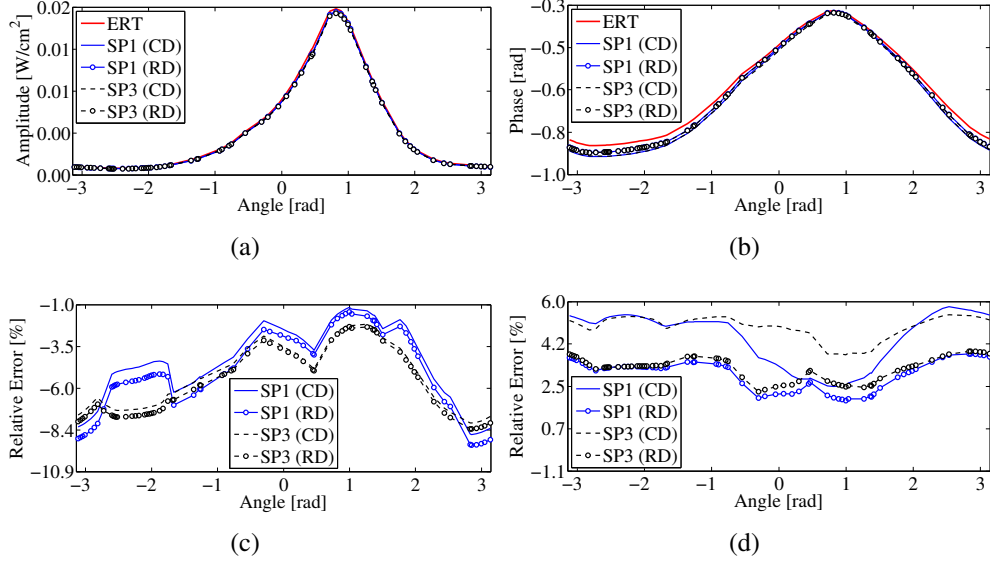


Figure 8.23: [Phantom 4] (a,b) Amplitude and phase of the fluence at mesh nodes within the internal cross-section. (c,d) Error in the amplitude and the phase of the fluence of the SP_N solutions relative to the benchmark ERT (S_{12}) solution.

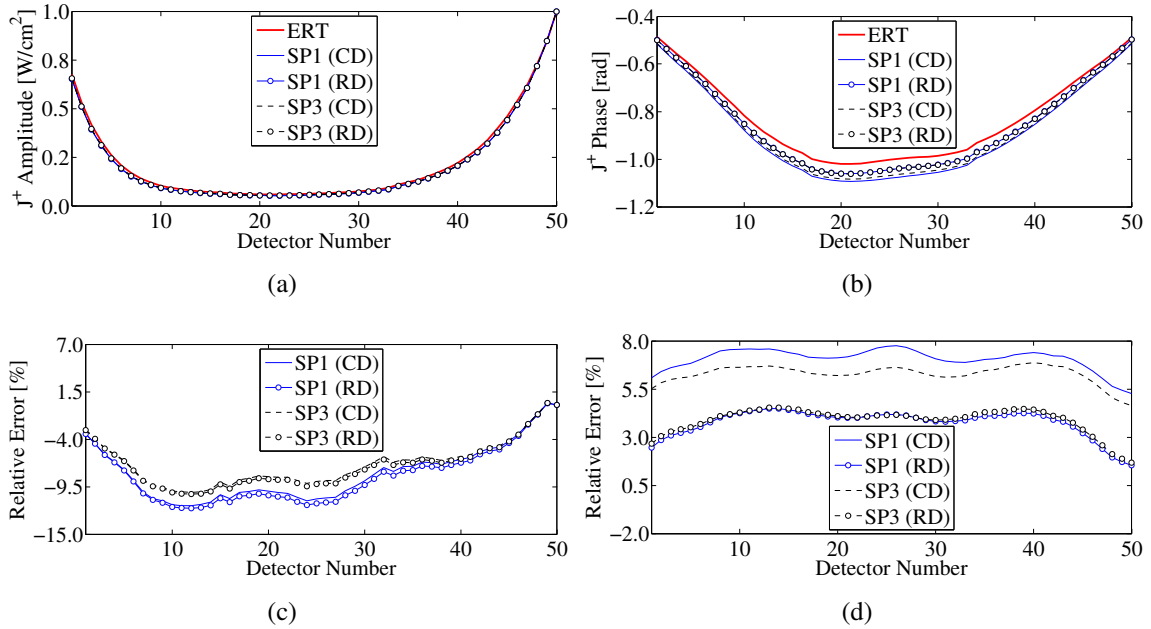


Figure 8.24: [Phantom 4] (a,b) Amplitude and phase of the partial current, $J^+(\mathbf{r})$. (c,d) Error in the amplitude and the phase of the partial current, $J_e^+(\mathbf{r})$, computed with the SP_N models relative to the benchmark ERT (S_{12}) solution.

Table 8.12: Computation time for phantom 4.

Model	Time [s]	RAM [MB]
SP ₁ (RD)	44.5	176.6
SP ₁ (CD)	70.7	194.3
SP ₃ (RD)	113.1	313.1
SP ₃ (CD)	155.1	359.4
ERT (S_{12})	6,715.4	6,140.0

8.5 Discussion

Overall, the SP₃ model is better at approximating the non-diffuse properties of light propagation throughout the entire phantom, and in particular at distances near the source. This is in accordance with our expectations, as it is well known that incoming photons do not become diffuse until they travel a short distance within media with absorbing and scattering properties. Thus, the SP₃ solution is able to better approximate the “transport” behavior of photons at short distances from the source. Typically, the first composite moment (φ_1) is a close approximation to the final fluence value. However, to account for “transport” properties, φ_1 is “corrected” by φ_2 , particularly near the source. It is this correction that generally leads to improved solutions with the SP₃ model.

The differences between the two SP_N models (CD and RD) are best observed when considering the phase of the fluence. In general, the SP_N (RD) models can more accurately approximate the phase of the fluence and the phase of the partial current than the SP_N (CD) models. In contrast, the SP_N (CD) models only marginally outperform the SP_N (RD) models in approximating the amplitudes of the fluence and the partial current.

We recommend using the SP_N (RD) model over the SP_N (CD) model. This recommendation is justified by overall performance of the two models; both models approximate the amplitude to similar degrees of accuracy. The SP_N (RD) model, however, is substantially better at approximating the phase values of the fluence and the partial current. In addition, the SP_N (RD) model requires lower computation time compared to the SP_N (CD) model. Furthermore, the SP₃ (RD) model is significantly better at approximating the benchmark solutions than the SP₁ (RD) model.

Thus, we specifically recommend using the SP_3 (RD) model as the light propagation model in DOT applications.

In Chapter 9 we present an algorithm for performing reconstructions of the absorption and the scattering coefficients using the SP_N (RD) model, hereafter referred to simply as the SP_N model.

Finite volume SP_N Model: Inverse Problem

The image reconstruction step in DOT refers to the process of solving an inverse problem to “re-construct” unknown tissue properties from the data at our disposal, such as photon counts measured with CCD cameras or optical fibers connected to photodiodes. Tissue properties of interest typically include μ_a and μ_s , but more generally, may include any variable that appears in the forward model. The variable or variables to reconstruct are denoted simply as μ . Due to the inability to directly invert the forward model, the reconstruction step is typically a large-scale optimization process, where the goal is to find the optical properties that minimize an objective function that measures the error between simulated detector readings generated with the forward model and the experimentally measured data.

That is, given our understanding of the light probing the tissue (O), we wish to determine the value and distribution of μ inside the medium that leads to the experimentally measured data M . Assuming a forward model $\mathcal{F}(\mu)$ that mathematically describes the way in which light propagates through tissue and can produce “simulated” detector measurements P , analogous to M , we can construct an objective function $f(\mu)$ that quantifies how well P , produced by $\mathcal{F}(\mu)$, compares to M . Then, the goal is to find the value of μ that minimizes $f(\mu)$ as this also minimizes the mismatch between P and M .

In this chapter we focus on presenting a summary of the necessary mathematics needed to implement an efficient DOT reconstruction algorithm for the FV-FD- SP_N model for reconstruction

of the absorption and the scattering coefficients. While the theory is presented in a general form that is well suited for application to any deterministic light propagation model, we focus on the implementation of the techniques presented in this chapter to the FD-SP_N model in (2.22), where the imaginary term arising from modulation of the light source does not appear in the diffusion-like coefficients.

In general, the objective function is the least-squares cost function, which minimizes the residual sum of the squared error. In its most general form, the objective function is written as

$$f(\mu, u) = \frac{1}{2} (M - P)^\top \overline{(M - P)}. \quad (9.1)$$

Here $(\overline{\cdot})$ represents the complex conjugate of (\cdot) and P is the projection of the solution u of the forward problem (i.e. fluence or radiance intensity) onto detectors, such that

$$P = Qu. \quad (9.2)$$

Q is a projection operator that maps the solution to the forward problem u onto boundary detectors as defined in Section 8.3.2. In this work, P is the partial current at the boundary, $J^+(r, \omega)$, where r denotes spatial position and ω is the modulation frequency of the source. The optimization problem necessary for finding the value of μ that minimizes (9.1) is typically written, in its most general form, as the following PDE-constrained quadratic problem,

$$\begin{aligned} \min \quad & f(\mu, u) = \frac{1}{2} (M - Qu)^\top \overline{(M - Qu)}, \\ \text{subject to} \quad & c(\mu, u) = A(\mu)u - b = 0. \end{aligned} \quad (9.3)$$

The operator matrix A and vector b represent the discretized version of the FD-SP_N model.

The problem in (9.3) is generally referred to as an “equality-constrained” optimization problem because it requires an accurate solution of the forward problem [i.e. $A(\mu)u - b$].

We explore two principal techniques for solving (9.3), the so called “unconstrained” and “PDE-constrained” optimization methods. The main difference between the techniques lies in how dependencies between μ and u are treated. The PDE-constrained method has been shown to yield significant improvements in computational efficiency compared to the unconstrained method. However this gain comes at a cost of increased algorithm complexity that results in implementation challenges.

With both methods, the goal is to solve a quadratic problem. In general, all quadratic problems require the computation of the first and second order derivatives of the objective function (9.1) with respect to all independent variables (i.e. μ for the unconstrained method; u and μ for the PDE-constrained method).¹ This procedure can be time consuming and impractical for optimization problems of even moderate size. Fortunately, numerical techniques for solving these types of quadratic problems have been widely studied in various areas of engineering and mathematics, and multiple efficient techniques for obtaining the inverse solution have been reported.

In this work, the challenges associated with computing the gradient of the objective function are overcome by making use of the adjoint equation to the forward problem, which allows calculation of the gradient at the computational cost of solving one forward problem.

We address the challenges of computing the Hessian matrix, H , through the use of the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm. L-BFGS is a type of quasi-Newton optimization algorithm specifically developed to address the computational burden associated with computing H of $f(\mu)$ by allowing us to instead use an approximate Hessian matrix B that can be easily computed explicitly. Furthermore, with the L-BFGS algorithm we can avoid having to explicitly compute B , instead only the product of B^{-1} and its associated vectors need to be stored and updated for the calculation of the descent (or search) direction.

¹Indeed, in the unconstrained formulation, the forward variable is a function of the inverse variable such that $u(\mu)$. In contrast, in the PDE-constrained formulation, both u and μ are independent variables. This topic is explored further in the remainder of this chapter.

Another key aspect of the approach we pursue for solving the inverse problem with the SP_N model is in how we simplify the Hessian matrix in the case of the PDE-constrained algorithm. Additional computational efficiency is gained by implementing a reduced Hessian sequential quadratic programming (rSQP) scheme, which allows for computationally efficient solutions to the optimization problem by reducing the size of the Hessian matrix needed to be stored and updated.

In Section 9.1 we briefly summarize the theoretical developments associated with quasi-Newton schemes for solving quadratic problems. Then, we present the unconstrained and constrained algorithms for solving the SP_N model in Sections 9.2 and 9.3. Algorithms for computing the partial derivative of the SP_N forward model matrix, A , with respect to the inverse variables μ_a and μ_s (an important part of the reconstruction algorithm) are presented in Section 9.4. The unconstrained and PDE-constrained algorithms are validated and their performance is analyzed in Section 9.5. Chapter 9 concludes with a discussion in Section 9.6. Common variables that appear throughout this chapter and that are central to the reconstruction process are summarized in Table 9.1.

Table 9.1: Definition of inverse problem variables.

$A \in \mathbb{C}^{m \times m}$	SP_N forward model operator matrix
$Q \in \mathbb{R}^{d \times m}$	Partial current operator
$H \in \mathbb{R}^{n \times n}$	Hessian matrix
$M \in \mathbb{C}^d$	Experimental measurement vector ($d = \#$ of total measurement points)
$P \in \mathbb{C}^d$	Predicted measurement vector ($d = \#$ of total measurement points)
$u \in \mathbb{C}^m$	SP_N forward solution (i.e. fluence)
$b \in \mathbb{C}^m$	SP_N forward model source vector
$\lambda \in \mathbb{C}^m$	Adjoint/Lagrangian variable
$\mu \in \mathbb{R}^n$	DOT inverse variable (e.g. μ_a or μ'_s)
$f(\mathbf{x}) \in \mathbb{R}$	Objective function
$\nabla f(\mathbf{x}) \in \mathbb{R}^n$	Gradient of objective function
$\nabla^2 f(\mathbf{x}) \in \mathbb{R}^{n \times n}$	Hessian of objective function
$\mathbf{x} \in \mathbb{R}^n$	General inverse variable
$\Delta \mathbf{x} \in \mathbb{R}^n$	Update direction of inverse variable
$\alpha \in \mathbb{R}$	Update step size
$\phi \in \mathbb{R}$	Merit function
$D\phi \in \mathbb{R}$	Directional derivative of merit function

9.1 Quasi-Newton methods

Newton's method is the best known procedure for computing the roots of an equation, and its applications in numerical optimization are well established [205, 206]. Extensive research has been dedicated to the generalization of the procedure to a wide range of systems of both linear and nonlinear equations. Newton's method has a rapid rate of convergence, typically quadratic, making it particularly appealing as it converges very rapidly once an iterate is sufficiently close to the root of the function [206]. However, application of Newton's method to large scale optimization problems is typically prohibitively expensive from a computing stand point. Instead, other methods that exhibit convergence properties similar to Newton's method have been extensively researched and reported. One of these methods, so called quasi-Newton, is merely an approximation to Newton's method.

In this chapter we present a brief review of quasi-Newton methods that will serve as a foundation to the optimization algorithms we apply to solve (9.3). An excellent reference for an in-depth presentation of methods for numerical optimization is the book "Numerical Optimization" by Nocedal and Wright [205].

Consider a general objective function that is a function of the general inverse variable \mathbf{x} , $f(\mathbf{x})$. The goal of the optimization problem is to find the roots of this function, where the solution achieves an extremum. In particular, we are interested in the value of \mathbf{x} that minimizes $f(\mathbf{x})$.

To understand the application of quasi-Newton methods to solving this type of optimization problem, we first derive Newton's method for the following optimization problem,

$$\min_{\mathbf{x}} \quad g(\mathbf{x}), \quad (9.4)$$

where $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth function of \mathbf{x} . The special meaning of this function will become clear momentarily. Then, with an assumption of differentiability, the optimal solution \mathbf{x} that minimizes $g(\mathbf{x})$ will satisfy the following condition,

$$\nabla_{\mathbf{x}} g(\mathbf{x}) = 0. \quad (9.5)$$

We can apply Newton's method to solve for roots of (9.5), which corresponds to the value of \mathbf{x} that minimizes (9.4). In particular, Newton's method tells us that for a function $g(\mathbf{x})$ that satisfies $\nabla g(\mathbf{x}) = 0$, beginning with an initial guess \mathbf{x}^0 , a sequence \mathbf{x}^k for $k > 0$ is built such that it approaches the optimal solution \mathbf{x} . The relationship can be stated as follows,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + (\nabla g(\mathbf{x}))^{-1} g(\mathbf{x}). \quad (9.6)$$

In this work $g(\mathbf{x})$ is the gradient of the DOT objective function $f(\mu, u)$ in (9.1) and (9.3) [i.e. $g(\mathbf{x}) = \nabla f(\mathbf{x})$]. Thus, iterative formula (9.6) can be written as follows,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + (\nabla^2 f(\mathbf{x}))^{-1} \nabla f(\mathbf{x}). \quad (9.7)$$

The focus is on computing the term $(\nabla^2 f(\mathbf{x}))^{-1} \nabla f(\mathbf{x})$. The first step is to expand f by a second order Taylor series at $\mathbf{x} + \Delta \mathbf{x}$,

$$f(\mathbf{x} + \Delta \mathbf{x}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^\top \nabla^2 f(\mathbf{x}) \Delta \mathbf{x}. \quad (9.8)$$

The function $f(\mathbf{x} + \Delta \mathbf{x})$ attains its minimum when its gradient with respect to $\Delta \mathbf{x}$ vanishes. Thus, differentiating both sides of the expression with respect to $\Delta \mathbf{x}$ and equating the results to zero,

$$\frac{\partial}{\partial \Delta \mathbf{x}} f(\mathbf{x} + \Delta \mathbf{x}) \approx \frac{\partial}{\partial \Delta \mathbf{x}} \left(f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^\top \nabla^2 f(\mathbf{x}) \Delta \mathbf{x} \right), \quad (9.9)$$

yields a standard quadratic equation for the Newton direction or $\Delta \mathbf{x}$,

$$\nabla^2 f(\mathbf{x}) \Delta \mathbf{x} = -\nabla f(\mathbf{x}). \quad (9.10)$$

We recognize the gradient (∇f) and second-order derivative or Hessian ($\nabla^2 f$) in the above equation. Thus, given the gradient and Hessian of the objective function, this expression gives us a formula for computing $\Delta \mathbf{x}$,

$$\Delta \mathbf{x} = -[\nabla^2 f(\mathbf{x})]^{-1} \nabla f(\mathbf{x}). \quad (9.11)$$

Optimization algorithms that use the Newton direction have been shown to have a fast convergence rate, typically quadratic. However, the computational demands of directly calculating $\nabla^2 f(\mathbf{x})$ or its inverse are challenging and often insurmountable. As a result, the Newton method is typically not directly used for large scale optimization problems. Instead, so-called quasi-Newton algorithms, which approximate the product term, $[\nabla^2 f(\mathbf{x})]^{-1} \nabla f(\mathbf{x})$, are preferred because of the reduced computational demand and their superlinear convergence properties. Two well-established and particularly efficient quasi-Newton algorithms are the L-BFGS and the symmetric-rank-one (SR1) algorithms.

Given the formula (9.11) for computing the Newton direction $\Delta \mathbf{x}$, the unknown variable \mathbf{x} can be updated iteratively with quasi-Newton algorithms, such as the L-BFGS and SR1. More generally, the iterative formula is stated as follows,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \Delta \mathbf{x}^k. \quad (9.12)$$

Central to this process is the computation of the Newton direction $\Delta \mathbf{x}^k$ and the step length α^k . The inclusion of α^k is to ensure that the change along the Newton direction results in a lower objective function value compared to the current iteration. The process of finding the Newton direction and the corresponding step length are crucial in the reconstruction process as they affect the convergence speed and the accuracy of the solution.

The initial guess vector may be homogeneous or inhomogeneous, the choice reflecting our *a priori* information on the distribution of optical properties inside the medium. For example, in a homogeneous initial guess the optical properties may be set to $\mu_a = 0.3 \text{ cm}^{-1}$ and $\mu'_s = 10.0 \text{ cm}^{-1}$. Then, given the initial guess μ^0 , the optimization algorithm seeks to find sequential approximations $\{\mu^k\}_0^\infty$ that continuously (with few exceptions) decrease the objective function (or a secondary measure termed the “merit” function).

The rest of this chapter is devoted to presenting a practical method for computing ∇f and $\nabla^2 f$ so that $\Delta \mathbf{x}$ can be obtained. Formulae for the gradient and Hessian can be derived by directly differentiating the objective function (9.1). We consider the unconstrained and PDE-constrained methodologies for reconstructing the optical properties using the FV-FD-SP_N model.

In Sections 9.2 and 9.3 we describe the necessary steps needed to compute $\Delta \mathbf{x}^k$ with the unconstrained and the PDE-constrained approaches, respectively. Both optimization methods rely on the L-BFGS algorithm, which does not require computation or storage of the full Hessian of the objective function. Instead, it uses an easy-to-compute approximation to $\nabla^2 f^k$ that is updated after each inverse iteration. This ensures that the most recent changes to the gradient (∇f^k) inform the most up-to-date approximation to the Hessian B^k . For reference, the L-BFGS algorithm is presented in Appendix B.1.

Two principal techniques exist for finding the update direction and step length, the line search

and trust region methods. In this work we consider two specific types of line search methodologies for computing α^k , each varying in the definition of the merit function. The methodologies differ according to the type of optimization algorithm. The line search routine used in the unconstrained algorithm is presented in Section 9.2.3, while the line search method for the PDE-constrained algorithm is presented in Section 9.3.4.

9.2 Unconstrained optimization

In the unconstrained formulation of the inverse problem in DOT, u is treated as *dependent* variables of an independent variable μ . The objective function that minimizes the sum of squared errors is therefore written as

$$\min_{\mu} \quad f(\mu) = \frac{1}{2} [M - Qu(\mu)]^T \overline{[M - Qu(\mu)]}. \quad (9.13)$$

The forward model is explicitly written as

$$A(\mu)u(\mu) = b. \quad (9.14)$$

Alternatively, the objective function can be expressed as

$$\min_{\mu} \quad f(\mu) = \frac{1}{2} [M - Q(A^{-1}(\mu)b)]^T \overline{[M - Q(A^{-1}(\mu)b)]}. \quad (9.15)$$

This is called *unconstrained* because the original constraint $A(\mu)u - b = 0$ no longer appears in the minimization problem and we are free to select any value for the inverse variable μ that minimizes the objective function. In general, additional constraints on the acceptable values of μ

can be enforced, such as $\mu \geq 0$. In this work we do not consider any specific constraints. However, the procedure detailed herein can be easily extended to circumstances where constraints on μ exist.

9.2.1 Gradient of objective function

From (9.11) we know that the gradient of the objective function $[\nabla_{\mu} f(\mu)]$ is needed by the L-BFGS algorithm in order to compute the Newton direction $\Delta\mu$. Directly differentiating (9.13) with respect to μ ,

$$\nabla_{\mu} f(\mu, u) = \frac{\partial f(\mu, u)}{\partial u} \frac{\partial u(\mu)}{\partial \mu} = \frac{\partial u}{\partial \mu} \frac{\partial f}{\partial u}, \quad (9.16)$$

yields the gradient. The second term on the right hand side of (9.16), the derivative of the objective function with respect to the forward solution u , is obtained by differentiating (9.1) with respect to u (see Appendix B.2 for a detailed derivation),

$$\frac{\partial f(\mu, u)}{\partial u} = Q^{\top} \overline{(Qu - M)}. \quad (9.17)$$

The first term on the right hand side of (9.16), the derivative of the forward variable with respect to the inverse variable μ , is obtained by differentiating (9.14) with respect to u and solving for the appropriate term (see Appendix B.3 for a detailed derivation), resulting in

$$\frac{\partial u(\mu)}{\partial \mu} = -\frac{\partial Au}{\partial \mu} (A^{\top})^{-1} \quad (9.18)$$

Using the results from (9.17-9.18) on (9.16) yields the formula for computing the gradient of the objective function,

$$\nabla_{\mu} f(\mu, u) = -\frac{\partial Au}{\partial \mu} (A^{\top})^{-1} \left[Q^{\top} \overline{(Qu - M)} \right]. \quad (9.19)$$

Given the value of the inverse variable μ^k at inverse iteration k , the forward model (A^k) , and its partial derivative with respect to μ are updated using μ^k . The u^k term is computed using the latest forward model, A^k . The terms Q and b do not vary across iterations. Then, all terms in (9.19) are well defined. We recognize that the term

$$(A^{\top})^{-1} \left[Q^{\top} \overline{(Qu - M)} \right] = \lambda \quad (9.20)$$

as part of a linear problem which we can re-write as

$$A^{\top} \lambda = Q^{\top} \overline{(Qu - M)}. \quad (9.21)$$

This is the adjoint equation to the forward model, where we have introduced the adjoint variable λ . We can solve this linear system for λ and use the solution in (9.19). Then, to obtain the gradient of the objective function $[\nabla_{\mu} f(\mu, u)]$ it only remains to compute the term $\frac{\partial Au}{\partial \mu}$. The adjoint equation technique is the most efficient method for computing the gradient of the objective function [20, 175], as the cost associated with computing the gradient is on the same order of magnitude as the cost of solving a forward problem.

Using the adjoint variable, we can re-write the gradient of the objective function into a version that can be used in the L-BFGS algorithm,

$$\nabla_{\mu} f(\mu, u) = - \left(u^{\top} \frac{\partial A}{\partial \mu} \lambda \right)_{\text{Re}}. \quad (9.22)$$

The derivation of the final form of $\nabla_{\mu} f(\mu, u)$ is presented in Appendix B.4 for reference. A subroutine for computing $\frac{\partial A}{\partial \mu}$ is given in Section 9.4.

9.2.2 Merit function

The merit function for the unconstrained problem is simply the objective function itself (9.15),

$$\phi_{\eta}^k(u^k, \mu^k) = \frac{1}{2} (Qu^k - M)^{\top} \overline{(Qu^k - M)}. \quad (9.23)$$

The line search routine requires the computation of the directional derivative of $\phi_{\eta}^k(u^k, \mu^k)$ along the descent direction $\Delta p = \Delta \mu$, which is given by

$$D\phi_{\eta}^k(\mu^k) = \left(u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k \right)_{\text{Re}}^{\top} \Delta \mu^k. \quad (9.24)$$

This term is simply the inner product of the gradient of the objective function and the descent direction.

9.2.3 Unconstrained algorithm

The final form of the unconstrained algorithm is presented in Algorithm 2. A subroutine for computing the step size α^k is presented in Algorithm 3. The algorithm exits when the objective function reaches a tolerance ϵ (typically $\epsilon = 1.0 \times 10^{-5}$) or reaches a maximum allowed number of inverse iterations K . The tolerance of the GMRES solver varies throughout the algorithm and is denoted by τ .

The reconstruction algorithm requires solving the forward and adjoint problems using the restarted GMRES algorithm with high degrees of tolerance within the outer loop (typically up to 1.0×10^{-12}). These steps are necessary to obtain accurate approximations to the gradient of the objective function, $\nabla_{\mu}^k f(\mu^k, u^k)$, which is an integral part of the L-BFGS algorithm that updates the Newton direction, $\Delta\mu^k$. There is an additional requirement to solve the forward problem accurately inside the line search routine. This step, repeated until a satisfactory α^k is found, is necessary to ensure sufficient decrease in the merit function over the previous inverse iteration.

The repetitive need to solve forward and adjoint problems within each inverse iteration is the primary reason for the extensive computation time associated with the unconstrained reconstruction technique. This deficiency is addressed in Section 9.3 through the use of a PDE-constrained optimization approach.

Algorithm 2 Unconstrained optimization for DOT with the FD-FV-SP_N model.

```

INVERSE  $\leftarrow$  true
 $k \leftarrow 0$ 
 $\mu^k \leftarrow$  Set initial guess (e.g.  $\mu_a$  and  $\mu_s$ )
 $H^k \leftarrow I$  Set the initial value of the Hessian matrix to the identity matrix
 $A^k \leftarrow A(\mu^k)$ 
 $u^k \leftarrow A^{k-1} b$  (using GMRES,  $\tau = 1.0 \times 10^{-12}$ )
while INVERSE do
   $\lambda^k \leftarrow \left( A^{k\top} \right)^{-1} Q^\top (Qu^k - M)$  (using GMRES,  $\tau = 1.0 \times 10^{-12}$ )
   $\nabla_{\mu}^k f(\mu^k, u^k) \leftarrow - \left( u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k \right)_{\text{Re}}$ 
   $\Delta\mu^k \leftarrow - \left( H_r^k \right)^{-1} \nabla_{\mu}^k f(\mu^k, u^k)$  (using L-BFGS)
  Line search: Compute  $\alpha^k$  (see Algorithm 3)
   $\mu^{k+1} \leftarrow \mu^k + \alpha^k \Delta\mu$ 
   $A^{k+1} \leftarrow A(\mu^{k+1})$ 
   $u^{k+1} \leftarrow \left( A^{k+1} \right)^{-1} b$  (using GMRES,  $\tau = 1.0 \times 10^{-12}$ )
   $l^{k+1} \leftarrow f(\mu^{k+1}, u^{k+1})$ 
  if ( $l^{k+1} < \epsilon$  or  $k \geq K$ ) then
    INVERSE  $\leftarrow$  false
  else
     $k \leftarrow k + 1$ 
  end if
end while

```

Algorithm 3 Line search subroutine for unconstrained optimization.

```
Compute  $\phi(0)$  using (9.23)
Compute  $D\phi(0)$  using (9.24)
 $LINESEARCH \leftarrow \mathbf{true}$ 
 $i \leftarrow 0$ 
 $\alpha_i \leftarrow 1$ 
while  $LINESEARCH$  do
   $\mu_{i+1} \leftarrow \mu_i + \alpha_i \Delta \mu^k$ 
   $A_{i+1} \leftarrow A(\mu_{i+1})$ 
   $u_{i+1} \leftarrow (A_{i+1})^{-1} b$  (using GMRES,  $\tau = 1.0 \times 10^{-12}$ )
  Compute  $\phi(\alpha_i)$  using (9.23)
  if ( $\phi(\alpha_i) \geq \phi(0) + c\alpha_i D\phi(0)$  or  $i \geq i_{\max}$ ) then
     $\alpha^k \leftarrow \alpha_i$ 
     $LINESEARCH \leftarrow \mathbf{false}$ 
  else
     $\alpha_{i+1} = \alpha_i \cdot \max \left( \min \left( 0.9, \frac{D\phi(0) \alpha_i}{\phi(0) + \alpha_i D\phi(0) - \phi(\alpha_i)} \right), 0.1 \right)$ 
     $i \leftarrow i + 1$ 
  end if
end while
```

9.3 PDE-Constrained optimization

A powerful method for accelerating the image reconstruction process, and indeed for solving large scale optimization problems, is known as “PDE-constrained” optimization. The *key* difference between “unconstrained” and PDE-constrained optimization approaches is that the forward and inverse variables are treated as independent variables in the latter technique. This assumption allows simultaneous solutions to the DOT problem with respect to both the forward and inverse variables, μ and u , respectively.

The importance of this property cannot be over-stated. Typically, the unconstrained problem requires that the forward solution, u , be computed after each inverse iteration (i.e. after every update to μ). This process can be very time consuming because the forward problem typically must be solved very accurately (i.e. to a GMRES tolerance of approximately $1.0\text{E} - 12$). In contrast, the PDE-constrained algorithm does not require that the forward problem be computed to such degree of accuracy. Instead, the value of u is iteratively updated in parallel to the update of μ . This results in significant reduction in computation time.

A second technique that is important in obtaining significant acceleration during the PDE-constrained optimization process involves making simplifications to the Hessian matrix of the quadratic problem that results from the DOT objective function. This method is commonly referred to as a reduced-Hessian or reduced sequential quadratic programming (rSQP) approach. The SQP approach itself is particularly effective at solving large-scale non-linear optimization problems [205]. The rSQP method improves on the SQP method by reducing the effective size of the Hessian matrix that needs to be computed (or approximated) and, therefore, provides significant gains in computation time.

In this section we present the necessary mathematics for understanding the rSQP PDE-constrained optimization approach. We specifically focus on the application of this technique to solving the DOT problem with the FD-FV-SP_N model.

We begin the derivation by reformulating (9.3), exploiting the fact that we now treat μ and u as independent variables. Using the Lagrange multiplier λ , the objective function is written, subject to the forward problem, as

$$\min_{\mu, u, \lambda} \quad \mathcal{L}(\mu, u, \lambda) = \frac{1}{2}(Qu - M)^\top \overline{(Qu - M)} + \lambda^\top (A(\mu)u - b). \quad (9.25)$$

This is a standard PDE-constrained problem which can be simultaneously solved for the forward and inverse solutions. For convenience, the problem can be written as

$$\min_{\mathbf{x}} \quad \mathcal{L}(\mathbf{x}) = f^I + \lambda^\top f^F, \quad (9.26)$$

where f^I denotes the inverse error and f^F denotes the error contribution from the forward model. The Lagrangian variable, λ , is a measure of the ratio between the inverse and forward errors. The goal of the optimization problem is to find the optimal values of μ , u , and λ that minimize $\mathcal{L}(\mathbf{x})$, where for notational purposes, we define the total unknown variable \mathbf{x} as

Table 9.2: Definition of PDE-constrained optimization terms.

$$f^I = \frac{1}{2}(Qu - M)^\top \overline{(Qu - M)}$$

$$f^F = Au - b$$

$$f_\mu^I = 0$$

$$f_{\mu\mu}^I = 0$$

$$f_u^I = Q^\top \overline{(Qu - M)}$$

$$f_{uu}^I = Q^\top Q$$

$$f_{\mu u}^I = 0$$

$$f_{u\mu}^I = 0$$

$$f_\mu^F = u^\top \frac{\partial A}{\partial \mu}$$

$$f_{\mu\mu}^F = u^\top \frac{\partial^2 A}{\partial \mu^2}$$

$$f_u^F = A^\top$$

$$f_{uu}^F = 0$$

$$f_{\mu u}^F = \frac{\partial A}{\partial \mu}$$

$$f_{u\mu}^F = \frac{\partial A}{\partial \mu}$$

$$f_\lambda^I = 0$$

$$f_{\lambda\lambda}^I = 0$$

$$f_{\mu\lambda}^I = 0$$

$$f_{\lambda\mu}^I = 0$$

$$f_{u\lambda}^I = 0$$

$$f_{\lambda u}^I = 0$$

$$W = \begin{bmatrix} f_{\mu\mu}^{F\top} & f_{\mu u}^{F\top} \\ \lambda^\top f_{u\mu}^F & \lambda^\top f_{uu}^F \end{bmatrix}$$

$$F = [f_\mu^F \quad f_u^F]$$

$$C = \begin{bmatrix} f_\mu^I \\ f_u^I \end{bmatrix}$$

$$\Delta p = \begin{bmatrix} \Delta \mu \\ \Delta u \end{bmatrix}$$

$$\hat{Z}^k = -(f_u^{F^k})^{-1} f^{F^k}$$

$$\hat{Y}^k = -(f_u^{F^k})^{-1} f_\mu^{F^k}$$

$$\mathbf{x} = \begin{bmatrix} \mu \\ u \\ \lambda \end{bmatrix} . \quad (9.27)$$

Then, applying the techniques from Section 9.1 to (9.26), yields a standard quadratic equation for $\Delta \mathbf{x}$ as a function of the gradient ($\nabla \mathcal{L}$) and second-order derivative or Hessian ($\nabla^2 \mathcal{L}$),

$$\nabla^2 \mathcal{L} (\mathbf{x}) \Delta \mathbf{x} = -\nabla \mathcal{L} (\mathbf{x}) . \quad (9.28)$$

Because this is an “all-in-one” updating scheme, finding $\Delta \mathbf{x}$ implies that we determine the Newton directions of the inverse variable ($\Delta \mu$), the forward variable (Δu), and the Langrange multiplier ($\Delta \lambda$). The rest of this chapter is devoted to presenting a practical method for computing $\nabla \mathcal{L}$ and $\nabla^2 \mathcal{L}$ so that $\Delta \mathbf{x}$ can be obtained.

Formulae for the gradient and Hessian can be derived by directly differentiating the objective function (9.26). Minimizing solutions can be obtained at points where the first order Karush-Kuhn-Tucker (KKT) conditions are met. That is, at points where the derivative of the functional $\mathcal{L} (\mathbf{x})$ with respect to \mathbf{x} (i.e. its arguments μ , u , and λ) vanishes or

$$\nabla_{\mathbf{x}} \mathcal{L} (\mathbf{x}) = \mathbf{0} . \quad (9.29)$$

An expression for the gradient of the Lagrangian function is obtained by differentiating the inverse f^I and forward f^F error functions with respect to \mathbf{x} .

$$\nabla \mathcal{L}(\mathbf{x}) = \begin{bmatrix} \nabla_{\mu} \mathcal{L}(\mathbf{x}) \\ \nabla_u \mathcal{L}(\mathbf{x}) \\ \nabla_{\lambda} \mathcal{L}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} f_{\mu}^I + f_{\mu}^{F\top} \lambda \\ f_u^I + f_u^{F\top} \lambda \\ f_{\lambda}^I + f^F \end{bmatrix} \quad (9.30)$$

The expression $\nabla_{\mu} \mathcal{L}(\mathbf{x})$ is the derivative of the forward operator with respect to the inverse variable μ (i.e. the sensitivity equation of the Lagrangian), while the term $\nabla_u \mathcal{L}(\mathbf{x})$ can be recognized as the adjoint equation to the forward problem. The expression for $\nabla_{\lambda} \mathcal{L}(\mathbf{x})$ is equivalent to the residual of the discretized SP_N forward model.

An expression for the Hessian matrix is obtained from the second order derivatives of the Lagrangian with respect to the forward, inverse, and Lagrange variables.

$$\nabla^2 \mathcal{L}(\mathbf{x}) = \begin{bmatrix} f_{\mu\mu}^I + f_{\mu\mu}^{F\top} \lambda & f_{\mu u}^I + f_{\mu u}^{F\top} \lambda & f_{\mu\lambda}^I + f_{\mu}^{F\top} \\ f_{u\mu}^I + \lambda^{\top} f_{u\mu}^F & f_{uu}^I + \lambda^{\top} f_{uu}^F & f_{u\lambda}^I + f_u^{F\top} \\ f_{\lambda\mu}^I + f_{\mu}^{F\top} & f_{\lambda u}^I + f_u^{F\top} & f_{\lambda\lambda}^I \end{bmatrix} \quad (9.31)$$

Remembering that u , μ , and λ are treated as independent variables, we can easily compute the first and second order derivatives of the forward and inverse error functions as summarized in Table 9.2. Using the definitions in Table 9.2, (9.30-9.31) can be simplified to the following equations.

$$\nabla \mathcal{L}(\mu) = \begin{bmatrix} f_{\mu}^{F\top} \lambda \\ f_u^I + f_u^{F\top} \lambda \\ f^F \end{bmatrix} \quad (9.32)$$

$$\nabla^2 \mathcal{L}(\mu) = \begin{bmatrix} f_{\mu\mu}^{F\top} \lambda & f_{\mu u}^{F\top} \lambda & f_{\mu}^{F\top} \\ \lambda^\top f_{u\mu}^F & \lambda^\top f_{uu}^F & f_u^{F\top} \\ f_{\mu}^{F\top} & f_u^{F\top} & 0 \end{bmatrix} \quad (9.33)$$

We can now re-write (9.28) with the expanded gradient and Hessian terms defined by (9.32-9.33), where we have introduced the variables F and C to denote the partial derivatives of the forward and inverse error functions, respectively (defined in Table 9.2).

$$\begin{bmatrix} W & F^\top \\ F & 0 \end{bmatrix} \begin{bmatrix} \Delta p \\ \Delta \lambda \end{bmatrix} = - \begin{bmatrix} C + F^\top \lambda \\ f \end{bmatrix} \quad (9.34)$$

This quadratic system can be solved using a quasi-Newton algorithm, such as L-BFGS, as we now have equations to compute the gradient of the objective function with respect to all unknown variables. However, solving for the optimal solution at this point can be difficult because of the burden associated with computing and updating the Hessian matrix, which is now larger than the Hessian matrix in the unconstrained problem. To overcome this challenge, various techniques have been widely reported that allow for the use of approximations to the Hessian matrix and that lead to significant reductions in its size and complexity.

9.3.1 Reduced Hessian approach

The system of equations defined by (9.34) can be written in quadratic form by considering the solution to the following equivalent system,

$$\underset{\Delta p, \lambda}{\text{minimize}} \quad \mathbb{L}(\Delta p, \lambda) = \left(\frac{1}{2} \Delta p^\top W \Delta p + \Delta p^\top C \right) + (\lambda^{k+1})^\top (F \Delta p + f), \quad (9.35)$$

where the relationship $\lambda^{k+1} = \lambda^k + \Delta\lambda^k$ is used. The equivalence to the original system is seen by differentiating $\mathbb{L}(\Delta p, \lambda)$ with respect to Δp (9.36) and λ (9.37), setting the expressions equal to zero, and then adding the resulting expressions. The resulting system of equations is equivalent to (9.34).

$$\frac{\partial}{\partial \Delta p} \mathbb{L} = 0 = W \Delta p + C + F^\top \lambda^{k+1} \quad (9.36)$$

$$\frac{\partial}{\partial \lambda} \mathbb{L} = 0 = F \Delta p + f \quad (9.37)$$

We begin by linearizing the inverse and forward variables. Note that we have separated the forward (u) and inverse (μ) variables from the Lagrange variable (λ). We have grouped the Δu and $\Delta \mu$ terms together into a new variable Δp , such that

$$\Delta p^k = Z^k + Y^k \Delta \mu^k. \quad (9.38)$$

The values of Z and Y have been studied in literature and are typically optimally defined as the following functions of f^I and f^F ,

$$\begin{aligned} Z^k &= \begin{bmatrix} \mathbf{0} \\ \hat{Z}^k \end{bmatrix}, \\ Y^k &= \begin{bmatrix} \mathbf{I} \\ \hat{Y}^k \end{bmatrix}. \end{aligned} \quad (9.39)$$

The terms \hat{Z}^k and \hat{Y}^k are defined by (9.40) and (9.41).

$$\hat{Z}^k = -(f_u^k)^{-1} f^k \quad (9.40)$$

$$\hat{Y}^k = -(f_u^k)^{-1} f_\mu^k \quad (9.41)$$

Therefore, substituting (9.38) and (9.39) into (9.36) and multiplying both sides by $Y^{k\top}$, results in

$$Y^{k\top} W^k Y^k \Delta\mu^k + Y^{k\top} W^k Z^k + Y^{k\top} F^{k\top} \lambda^{k+1} = -Y^{k\top} C^k. \quad (9.42)$$

Which can be simplified to

$$H_r^k \Delta\mu^k + d_r^k = -g_r^k. \quad (9.43)$$

The term $H_r^k = Y^{k\top} W^k Y^k$ is the “reduced Hessian,” $d_r^k = Y^{k\top} W^k Z^k$ is the “cross-term,” $Y^{k\top} F^{k\top} = 0$, and the “reduced gradient” is given by

$$g_r^k = Y^{k\top} C^k. \quad (9.44)$$

The update direction, $\Delta\mu^k$, to the inverse variable, μ^k , can be obtained from (9.43), after ignoring the cross term d_r^k , as follows,

$$\Delta\mu^k = -(H_r^k)^{-1} (g_r^k). \quad (9.45)$$

The L-BFGS algorithm is used to compute and directly update the inverse of the reduced Hessian matrix $(H_r^k)^{-1}$ at each inverse iteration k . We can now solve for Δu using the second equation in

(9.34), where $F^k \Delta p^k = -f^k$ can be re-written as $f_\mu^k \Delta \mu^k + f_u^k \Delta u^k = -f^k$, and then solved and simplified as follows,

$$\begin{aligned}\Delta u^k &= - (f_u^k)^{-1} f^k - (f_u^k)^{-1} f_\mu^k \Delta \mu, \\ &= \hat{Z}^k - \hat{Y}^k \Delta \mu^k.\end{aligned}\tag{9.46}$$

Expressions (9.40) and (9.41) were used to simplify (9.46). We can re-write the original quadratic problem, where the Hessian (W) is reduced to H_r^k at the k^{th} inverse iterations, as follows,

$$\begin{bmatrix} H_r^k & 0 & f_\mu^\top \\ 0 & Q^\top Q & f_u^\top \\ f_\mu & f_u & 0 \end{bmatrix} \begin{bmatrix} \Delta \mu \\ \Delta u \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} c_\mu + f_\mu^\top \lambda \\ c_u + f_u^\top \lambda \\ f \end{bmatrix}.\tag{9.47}$$

We could derive an expression for the update direction of the Lagrange multiplier $\Delta \lambda$, however, instead we compute the updated value of the variable λ^{k+1} using the adjoint equation method. Differentiating the Lagrangian function in (9.25) with respect to the forward variable u and setting the derivative to zero results in,

$$0 = -Q^\top \overline{(Qu - M)} + A^\top \lambda.\tag{9.48}$$

Then, λ^{k+1} can be obtained by solving the following problem with a linear solver, such as GMRES.

$$\lambda^{k+1} = - \left(A (\mu^{k+1})^\top \right)^{-1} \left[Q^\top \overline{(Qu^{k+1} - M)} \right]\tag{9.49}$$

In this work, the update to the Lagrangian variable λ^{k+1} is always computed by solving the adjoint problem.

9.3.2 Summary of formulations

Finally, we have well-prescribed sets of equations for computing the update to each of the three optimization variables. The forward and inverse variables are updated along Δu and $\Delta \mu$, such that

$$u^{k+1} = u^k + \alpha^k \Delta u^k \quad (9.50)$$

$$\mu^{k+1} = \mu^k + \alpha^k \Delta \mu^k \quad (9.51)$$

where $\Delta \mu$ and Δu are given by (9.45) and (9.46), respectively. The update to the Lagrange variable λ is obtained by solving the adjoint problem to the forward model as specified by (9.49), reproduced below,

$$\lambda^k = -\left(A(\mu^k)^\top\right)^{-1} \left[Q^\top(Qu^k - M)\right]. \quad (9.52)$$

The adjoint equation (9.52) is solved using the restarted GMRES solver, with the right hand side vector inside brackets, $[\cdot]$. Then, given λ^k and by expanding the expression for the reduced gradient (9.44) into (9.45) we can obtain a clear expression for $\Delta \mu$ that is a function of the forward model, forward solution, and the Lagrangian variable. The reduced gradient is simplified to

$$g_r^k = -\left(u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k\right)_{\text{Re}}. \quad (9.53)$$

The details of this expansion are given in Appendix B.5. Given the gradient of the objective function with respect to the inverse variable μ , the update direction $\Delta \mu$ is simply given by

$$\Delta\mu^k = -(H_r^k)^{-1} \left(u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k \right)_{\text{Re}}. \quad (9.54)$$

This vector, the Newton direction, is computed using the L-BFGS algorithm. The update to the forward variable Δu can be simplified by expanding (9.46), which results in

$$\Delta u^k = - (A^k)^{-1} \left[(A^k u^k - b) + u^{k\top} \frac{\partial A^k}{\partial \mu} \Delta \mu^k \right]. \quad (9.55)$$

Thus, to obtain the Newton direction (update direction) of the forward variable, Δu^k , the system of linear equations in (9.55) has to be solved using the restarted GMRES algorithm, where the right hand side vector is the bracketed term, $[\cdot]$. See Appendix B.6 for a detailed derivation of (9.55).

Now, the Lagrange variable λ (9.52), inverse variable μ (9.54), and forward variable u (9.55) are all functions of the SP_N forward model matrix A and source vector b . The optimization problem and its associated variables are, therefore, all defined.

9.3.3 Merit function

The merit function is designed to ensure that the error contribution from the inverse and forward variables both decrease. The objective function used in this work is the following,

$$\phi_\eta^k(u^k, \mu^k) = \frac{1}{2} (Qu^k - M)^\top \overline{(Qu^k - M)} + \eta_k \|A^k u^k - b^k\|_1. \quad (9.56)$$

The Line search routine requires the computation of the directional derivative of $\phi_\eta^k(u^k, \mu^k)$ along the descent direction $\Delta p = (\Delta u, \Delta \mu)$, which is given by

$$D\phi_{\eta}^k(u^k, \mu^k) = \left(u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k \right)_{\text{Re}}^{\top} \Delta \mu^k - \eta_k \|A^k u^k - b\|_1. \quad (9.57)$$

The derivation of $D\phi_{\eta}^k(u^k, \mu^k)$ is given in Appendix B.7.

9.3.4 rSQP PDE-constrained algorithm

A general algorithm for iteratively solving for the optimal values of μ and u that simultaneously minimizes the error contribution from the inverse and forward models is presented in Algorithm 4. The algorithm exits when the objective function reaches a tolerance ϵ (typically $\epsilon = 1.0 \times 10^{-5}$) or reaches a maximum allowed number of inverse iterations K . The tolerance of the GMRES solver varies throughout the algorithm and is denoted by τ .

In contrast to the unconstrained reconstruction algorithm, the PDE-constrained algorithm does not require the solution of the forward and the adjoint problems with accurate tolerances. Instead, the forward problem is solved accurately only once, at the beginning of the inverse algorithm ($\tau = 1.0 \times 10^{-8}$). The adjoint problem is solved with relaxed tolerances compared to the unconstrained problem. There is significant freedom in choosing this tolerance, but we have found that an adequate range is between $\tau = 1.0 \times 10^{-4}$ and $\tau = 1.0 \times 10^{-6}$. Solving the adjoint problem with more accuracy may only lead to marginal increases in accuracy, but it can result in a significant increase in computation time.

The Newton direction for the forward variable, Δu^k , is obtained by solving a linear problem with the GRMES algorithm, but only a very relaxed tolerance is required. We have found that an adequate range is between $\tau = 1.0 \times 10^{-2}$ and $\tau = 1.0 \times 10^{-4}$.

The line search process is repeated until a satisfactory α^k is found, which ensures a sufficient decrease in the merit function over the previous inverse iteration. However, the line search routine of the PDE-constrained algorithm does not require solving a linear problem, unlike the line search routine in the unconstrained algorithm. This is the main reasons why the PDE-constrained

reconstruction algorithm achieves convergence at reduced computation time compared to the unconstrained algorithm.

Algorithm 4 rSQP PDE-constrained optimization for DOT with the FD-FV-SP_N model.

```

INVERSE ← true
k ← 0
 $\mu^k \leftarrow$  Set initial guess (e.g.  $\mu_a$  and  $\mu_s$ )
 $A^k \leftarrow A(\mu^k)$ 
 $u^k \leftarrow$  Compute from  $A^{k\top} u^k = b$  using GMRES ( $\tau = 1.0 \times 10^{-8}$ )
 $H^k \leftarrow I$  Set the initial value of the Hessian matrix to the identity matrix
while INVERSE do
     $\lambda^k \leftarrow$  Compute from  $A^k \lambda^k = Q^\top (Qu^k - M)$  using GMRES ( $\tau = 1.0 \times 10^{-4}$ )
     $g_r^k \leftarrow - \left( u^{k\top} \frac{\partial}{\partial \mu} A^k \right)^\top \lambda^k$ 
     $\Delta \mu^k \leftarrow - (H_r^k)^{-1} \left( u^{k\top} \frac{\partial}{\partial \mu} A^k \right)^\top \lambda^k$  using L-BFGS algorithm
     $\Delta u^k \leftarrow - (A^k)^{-1} \left[ (Au - b)^k + \left( u^\top \frac{\partial}{\partial \mu} A \right)^k \Delta \mu^k \right]$  using GMRES solver ( $\tau = 1.0 \times 10^{-2}$ )
    Line search: Compute  $\alpha^k$  (see Algorithm 5)
     $\mu^{k+1} \leftarrow \mu^k + \alpha^k \Delta \mu$ 
     $u^{k+1} \leftarrow u^k + \alpha^k \Delta u$ 
     $l^{k+1} \leftarrow f(\mu^{k+1}, u^{k+1}, \lambda^{k+1})$ 
    if ( $l^{k+1} < \epsilon$  or  $k \geq K$ ) then
        INVERSE ← false
    else
         $k \leftarrow k + 1$ 
    end if
end while

```

Algorithm 5 Line search subroutine for rSQP PDE-constrained algorithm.

```

Compute  $\phi(0)$  using (9.56)
Compute  $D\phi(0)$  using (9.57)
 $LINESEARCH \leftarrow \mathbf{true}$ 
 $i \leftarrow 0$ 
 $\alpha_i \leftarrow 1$ 
while  $LINESEARCH$  do
   $\mu_{i+1} \leftarrow \mu_i + \alpha_i \Delta \mu^k$ 
   $u_{i+1} \leftarrow u_i + \alpha_i \Delta u^k$ 
   $A_{i+1} \leftarrow A(\mu_{i+1})$ 
  Compute  $\phi(\alpha_i)$  using (9.56)
  if ( $\phi(\alpha_i) \geq \phi(0) + c\alpha_i\phi'(0)$  or  $i \geq i_{\max}$ ) then
     $\alpha^k \leftarrow \alpha_i$ 
     $LINESEARCH \leftarrow \mathbf{false}$ 
  else
     $\alpha_{i+1} = \alpha_i \cdot \max \left( \min \left( 0.9, \frac{D\phi(0) \alpha_i}{\phi(0) + \alpha_i D\phi(0) - \phi(\alpha_i)} \right), 0.1 \right)$ 
     $i \leftarrow i + 1$ 
  end if
end while

```

9.4 Gradient of forward model

The gradient of the forward model matrix A with respect to the inverse variables (i.e. optical properties) appears in the constrained and unconstrained optimization approaches. In this section we detail the derivation of formulae for computing the gradient with respect to μ_a and μ_s , as these are the two primary optical properties of interest in DOT.

It is common in DOT to want to avoid the direct computation of the partial derivative of A with respect to μ_s (i.e. $\frac{\partial A}{\partial \mu_s}$). Instead of μ_s , the reduced scattering coefficient $\mu'_s = (1 - g)\mu_s$ is preferred for simplicity. To further simplify the computation of $\frac{\partial A}{\partial \mu_s}$, the following additional approximation is made:

$$\mathfrak{D} = \frac{1}{3(\mu_a + \mu'_s)}, \quad (9.58)$$

$$\approx \frac{1}{3\mu'_s}. \quad (9.59)$$

Then, instead of solving for μ_s or μ'_s , one typically solves for \mathfrak{D} (i.e. the diffusion coefficient), from which the scattering coefficient can be recovered. This approximation only works well for applications in which the diffusion model is valid (i.e. $\mu_a \ll \mu'_s$).

In this work we do not make these approximations. Instead, we differentiate A with respect to μ_a and μ_s directly. The formulae for computing this term depends on the order of the SP_N model. Algorithms 6 and 7 present the definitions of the partial derivatives of A with respect to μ_s and μ_a for the SP_3 model, respectively.

Algorithm 6 Gradient of SP_3 model matrix A with respect to μ_s .

```

 $\nabla_{\mu_s} \mathbf{A} \leftarrow 0$ 
for ( $i = 0; i < M; i++$ ) do
   $\nabla_{\mu_s} \mathbf{A}_{ii} \leftarrow \Delta V_i$ 
  for ( $j = 0; j < S_T; j++$ ) do
    if ( $S_j$  is interior surface) then
       $k \leftarrow$  neighbor node number
       $\beta_1 \leftarrow \beta_1 - \frac{3(1-g)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_i^2$ 
       $\beta_2 \leftarrow \beta_2 - \frac{7(1-g^3)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_i^2$ 
       $\nabla_{\mu_s} \mathbf{A}_{iik} \leftarrow -\frac{3(1-g)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_k^2$ 
       $\nabla_{\mu_s} \mathbf{A}_{iki} \leftarrow \frac{3(1-g)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_i^2$ 
       $\nabla_{\mu_s} \mathbf{A}_{ikk} \leftarrow \frac{3(1-g)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_k^2$ 
       $\nabla_{\mu_s} \mathbf{A}_{i+M,i+M,k} \leftarrow -\frac{7(1-g^3)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_k^2$ 
       $\nabla_{\mu_s} \mathbf{A}_{i+M,k+M,i} \leftarrow \frac{7(1-g^3)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_i^2$ 
       $\nabla_{\mu_s} \mathbf{A}_{i+M,k+M,k} \leftarrow \frac{7(1-g^3)}{2l_s} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_k^2$ 
    end if
  end for
   $\nabla_{\mu_s} \mathbf{A}_{iii} \leftarrow \beta_1$ 
   $\nabla_{\mu_s} \mathbf{A}_{i+M,i+M,i} \leftarrow \frac{5(1-g^2)}{9l_s} [\Delta V]_i + \beta_2$ 
end for

```

Algorithm 7 Gradient of SP_3 model matrix A with respect to μ_a .

```

 $\nabla_{\mu_a} \mathbf{A} \leftarrow 0$ 
for ( $i = 0; i < M; i++$ ) do
   $\nabla_{\mu_a} \mathbf{A}_{ii} += \Delta V_i$ 
  for ( $j = 0; j < S_T; j++$ ) do
    if ( $S_j$  is interior surface) then
       $k \leftarrow$  neighbor node number
       $\beta_1 \leftarrow \beta_1 - \frac{3}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_i^2$ 
       $\beta_2 \leftarrow \beta_2 - \frac{7}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_i^2$ 
       $\nabla_{\mu_a} \mathbf{A}_{iik} \leftarrow -\frac{3}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_k^2$ 
       $\nabla_{\mu_a} \mathbf{A}_{iki} \leftarrow \frac{3}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_i^2$ 
       $\nabla_{\mu_a} \mathbf{A}_{ikk} \leftarrow \frac{3}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_1]_k^2$ 
       $\nabla_{\mu_a} \mathbf{A}_{i+M,i+M,k} \leftarrow -\frac{7}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_k^2$ 
       $\nabla_{\mu_a} \mathbf{A}_{i+M,k+M,i} \leftarrow \frac{7}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_i^2$ 
       $\nabla_{\mu_a} \mathbf{A}_{i+M,k+M,k} \leftarrow \frac{7}{2l_a} \left[ \frac{dA_j}{dr_j} \right] [\mathfrak{D}_2]_k^2$ 
    end if
  end for
   $\nabla_{\mu_a} \mathbf{A}_{iii} \leftarrow \frac{[\Delta V]_i}{l_a} + \beta_1$ 
   $\nabla_{\mu_a} \mathbf{A}_{i,i+M,i} \leftarrow -\frac{2}{3} \frac{[\Delta V]_i}{l_a}$ 
   $\nabla_{\mu_a} \mathbf{A}_{i+M,i,i} \leftarrow -\frac{2}{3} \frac{[\Delta V]_i}{l_a}$ 
   $\nabla_{\mu_a} \mathbf{A}_{i+M,i+M,i} \leftarrow \frac{[\Delta V]_i}{l_a} + \beta_2$ 
end for

```

9.5 Validation and numerical simulations

We validate the unconstrained and the PDE-constrained algorithms through numerical studies, using numerical phantoms with known optical properties. We test multiple ranges of optical properties to better understand how well each algorithm performs in terms of accuracy and computation speed.

9.5.1 Numerical phantoms

We compare the performance of the algorithms through simulations on a two-dimensional circular phantom and a three-dimensional phantom that resembles a human finger. Absorbing and scattering inclusions are placed inside each phantom and simulated measurement data (M) is generated using the SP_3 model on a dense FVM mesh and corrupted with 1.0% Gaussian white noise.² These two precautions help mitigate the risk of committing an inverse crime during the reconstruction process. We then attempt to reconstruct the location and optical properties of the phantom, beginning with a standard homogeneous initial guess for μ_a and μ'_s . All simulations on these phantoms are performed with $g = 0.95$, $\omega = 600$ MHz, and $n_m = 1.4$. These values are chosen to approximate the optical properties encountered in real-world applications.

The two-dimensional phantom is defined to have 8 distinct boundary sources distributed along the perimeter of a circle with a diameter of 2 cm (Fig. 9.1a). For each source a total of 110 detectors are defined along the perimeter of the phantom. The detectors are required to be at least 0.5 cm away from the source (Fig. 9.1b). In certain circumstances it is convenient to analyze the optical properties along a specific cross-section inside the medium. For this reason, a circular cross-section 0.5 cm from the center of the phantom is defined and will be referred to in subsequent sections (Fig. 9.1b).

²The two-dimensional mesh with which the simulated measurements are computed contains 37,247 nodes (73,236 tetrahedral elements), while the mesh on which the reconstructions are performed has 1,309 nodes (2,360 tetrahedral elements).

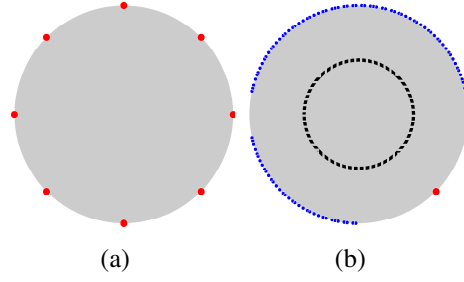


Figure 9.1: (a) Location of boundary sources (red). (b) Location of boundary detectors (blue) for a single boundary source (red) and placement of the internal cross-section of interest (black).

9.5.2 Performance evaluation

The accuracy of each reconstruction is quantified using the correlation coefficient, defined as

$$r = \frac{\sum_{i=1}^n (\phi_i^r - \overline{\phi_i^r}) (\phi_i^a - \overline{\phi_i^a})}{\sqrt{\sum_{i=1}^n (\phi_i^r - \overline{\phi_i^r})^2} \sqrt{\sum_{i=1}^n (\phi_i^a - \overline{\phi_i^a})^2}}, \quad (9.60)$$

where ϕ^a is the analytical (i.e. benchmark) solution and ϕ^r is the reconstruction image. The mean of each image is denoted by $(\overline{\cdot})$ and n refers to the total number of distinct mesh nodes. The analytical solution is known because it is used by the forward model to generate the “measurement” data M . The correlation coefficient r is between -1 and 1, with $r = 1.0$ corresponding to an exact match between the analytical and reconstruction images.

The computation efficiency is captured by plotting the value of the inverse error over time for each combination of SP_N model and reconstruction algorithm. The algorithm is set to exit when the inverse error decreases by four orders of magnitude from its original value (i.e. relative decrease of $1e-4$). We note that the accuracy of the reconstruction can be significantly improved by allowing the reconstruction process to continue for a longer amount of time. In this analysis, however, we only require the inverse error to decrease by four orders of magnitude from its original value (i.e. the final inverse error is 0.01% of the original error).

9.5.3 Disk phantom: case 1

The optical properties of the first phantom are well within the diffuse regime, where the diffusion model performs well as a light propagation model (Fig. 9.2a,f). The optical properties of the background medium are $\mu_a = 0.01 \text{ cm}^{-1}$ and $\mu'_s = 12.50 \text{ cm}^{-1}$; the absorbing inclusions have a value of $\mu_a = 0.05 \text{ cm}^{-1}$, while the low scattering inclusion is defined by $\mu'_s = 10.0 \text{ cm}^{-1}$. Reconstruction images of μ_a and μ'_s are presented in Fig. 9.2. In general, reconstructions obtained with the SP_3 model are more accurate than the reconstruction obtained with the SP_1 model. All SP_N models correctly identify the general location of each inclusion. However, reconstructions with the SP_1 model have significant boundary artifacts in both μ_a and μ'_s , and strong cross talk in μ_a images.

Analysis of the reconstruction values along the internal cross section allows for better understanding of the performance of the SP_N models (Fig. 9.3). Both models underestimate μ_a in the region of the absorbing inclusions and marginally underestimate μ_a in the region of the μ_s inclusion. In the case of μ'_s reconstructions, the SP_1 model severely underestimates μ'_s in the location of the scattering inclusion, while the SP_3 model accurately predicts the lower bounds of μ'_s . The cross-talk between absorption and scattering in the μ'_s reconstruction is evident in the SP_1 recon-

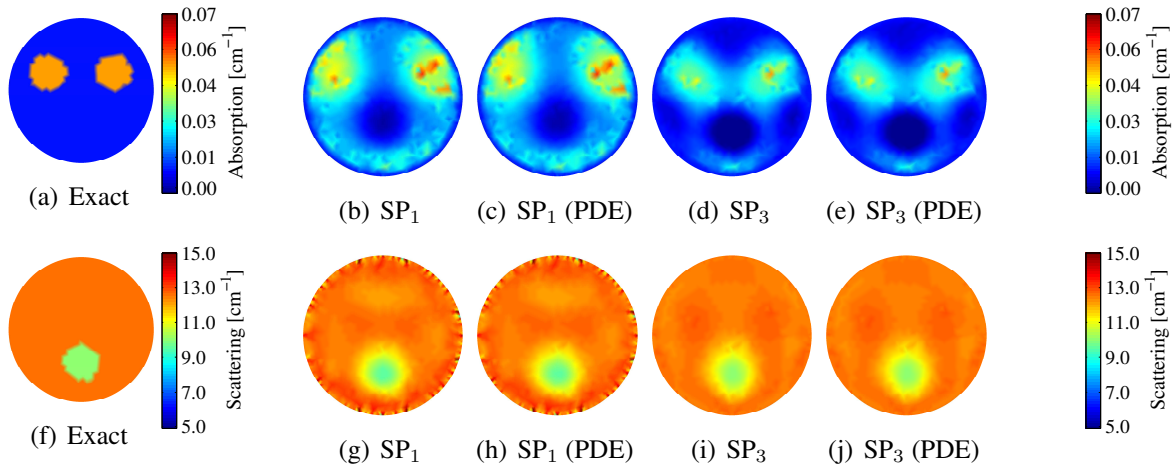


Figure 9.2: (Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).

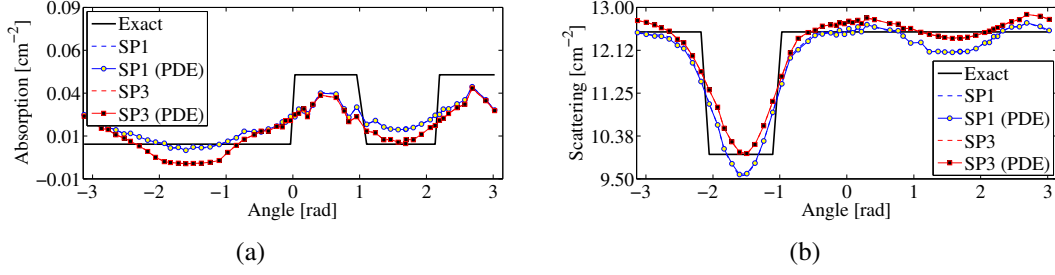


Figure 9.3: Values of (a) absorption and (b) scattering coefficient along the internal cross section defined in Fig. 9.1b.

struction, where the SP_1 models significantly underestimate μ'_s at the location of the absorbing inclusions. Thus, the SP_3 model is more resistant to cross talk between μ_a and μ_s .

The correlation coefficient between the reconstruction and the analytical solutions verify that the SP_3 model generates reconstructions that are significantly more accurate than reconstructions with the SP_1 model (Table 9.3) in both optical parameters, as is evident from the reconstruction images. With the SP_3 model, the μ'_s reconstruction is more accurate than the μ_a reconstruction, with correlation coefficients of $r = 0.85$ and $r = 0.66$, respectively. Similar patterns are observed for the SP_1 model, although the overall accuracy of the reconstructions are significantly lower than those obtained with the SP_3 model. The choice of optimization strategy, whether constrained or unconstrained, does not impact the overall accuracy of the reconstruction solution.

The computational advantages of using the PDE-constrained algorithm over the unconstrained algorithm is clear from Fig. 9.4a, where we plot the value of the objective function over time for each of the four reconstruction methods. In the case of the SP_1 model, the PDE-constrained algorithm converges approximately 12.2 times faster than the unconstrained algorithm. In the case of the SP_3 model, a speedup factor of 11.75 was achieved with the PDE-constrained algorithm

Table 9.3: Correlation coefficient between reconstruction and analytical solutions for the first phantom.

	SP_1	SP_1 (PDE)	SP_3	SP_3 (PDE)
$c(\mu_a)$ [%]	0.51	0.51	0.66	0.66
$c(\mu'_s)$ [%]	0.51	0.51	0.85	0.85

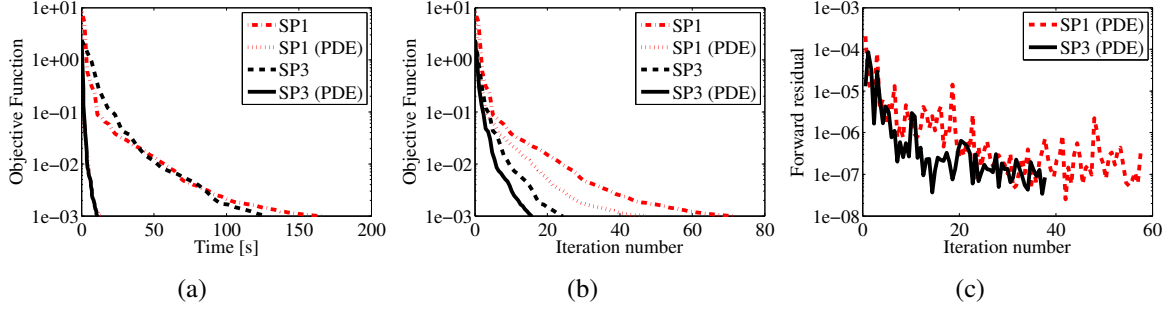


Figure 9.4: Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.

compared to the unconstrained algorithm.

The decrease in the inverse error after each inverse iteration is not significantly affected by the type of optimization algorithm, but is significantly affected by the choice of SP_N model (Fig. 9.4b). The SP_3 model exhibits a faster decay rate in the inverse error compared to the SP_1 model. Fig. 9.4c is an example of how the “forward” error decreases over inverse iterations. The forward error decreases in parallel to the inverse error. This phenomena is a fundamental property of the PDE-constrained algorithm.

9.5.4 Disk phantom: case 2

The optical properties of the second phantom are in the regime where the diffusion model is increasingly a poor approximation, with highly absorbing regions ($\mu_a = 0.5 \text{ cm}^{-1}$) and an area with low scattering values ($\mu'_s = 2.5 \text{ cm}^{-1}$) (Fig. 9.5a,f). The background properties are $\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu'_s = 10.0 \text{ cm}^{-1}$. These values closely resemble absorption and scattering of NIR light in various biological media, including the human breast ($\mu_a \in [0.34, 0.8] \text{ cm}^{-1}$, $\mu_s \in [181.0, 492.0] \text{ cm}^{-1}$, $g \in [0.95, 0.98]$) and human brain gray matter ($\mu_a \in [0.16, 0.49] \text{ cm}^{-1}$ and $\mu'_s \in [5.9, 9.3] \text{ cm}^{-1}$) [207]. Reconstructions of μ_a and μ'_s are shown in Fig. 9.5. As in the case of the diffuse phantom (Section 9.5.3), the reconstructions obtained with the SP_3 model are more accurate than those obtained with the SP_1 model.

The μ_a objects are more accurately resolved (based on separation) by the SP_3 model compared

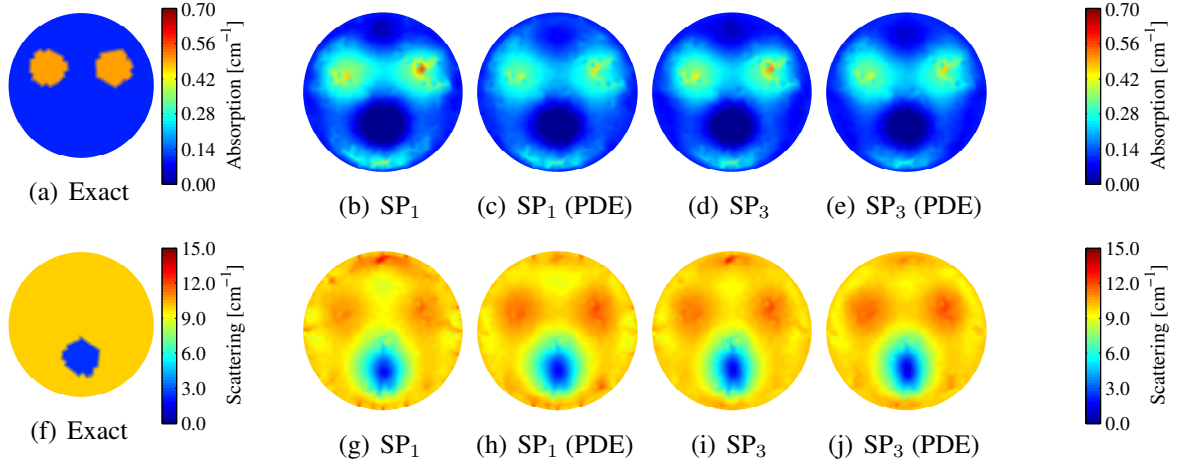


Figure 9.5: (Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).

to the SP_1 model. In the case of μ'_s , the reconstructions with the SP_1 model have more significant boundary artifacts compared to the reconstructions with the SP_3 model.

In general, the SP_3 model yields more accurate reconstructions than the SP_1 model, while reconstructions obtained with the PDE-constrained algorithm are as accurate as those obtained with the unconstrained algorithm (Table 9.4.). With the PDE-constrained algorithm, the correlation coefficient of the SP_3 model for μ_a is 0.68, while it is only 0.63 for the SP_1 model. In the case of μ'_s , the correlation coefficients are 0.82 and 0.79, for the SP_3 and SP_1 models, respectively. As in the case of the first phantom, the scattering coefficient reconstruction is more accurate than the absorption reconstruction.

The PDE-constrained algorithm achieves speedup factors of 16.6 and 10.7 compared to the unconstrained algorithm, using the SP_1 and SP_3 models, respectively (Fig. 9.6). The difference

Table 9.4: Correlation coefficient between reconstruction and analytical solutions for the second phantom.

	SP_1	SP_1 (PDE)	SP_3	SP_3 (PDE)
$c(\mu_a)$ [%]	0.64	0.64	0.68	0.68
$c(\mu'_s)$ [%]	0.79	0.80	0.82	0.82

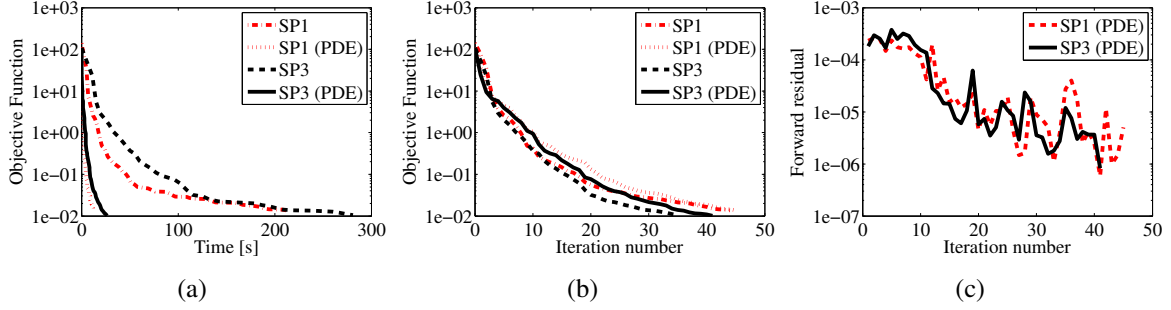


Figure 9.6: Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.

in computation time between the SP_1 and SP_3 models is surprisingly minimal (Fig. 9.6a), suggesting that the traditional computation superiority of the SP_1 model over higher order models is diminished when considering optical properties outside the diffuse regime. The similarities in computation performance between then SP_1 and SP_3 models are evident by considering the value of the objective function over inverse iterations (Fig. 9.6b) and the value of the forward model error over inverse iterations (Fig. 9.6c); the performance of the SP_3 is almost identical to that of the SP_1 model.

9.5.5 Disk phantom: case 3

The optical properties of the third phantom are chosen to be well outside the diffuse regime, with highly absorbing inclusions ($\mu_a = 1.0 \text{ cm}^{-1}$) and a region with low scattering values ($\mu'_s = 2.5 \text{ cm}^{-1}$), all inside a moderately diffuse background ($\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu'_s = 10.0 \text{ cm}^{-1}$) (Fig. 9.7a,f). These absorption and scattering values closely approximate the properties of most biological media (a comprehensive list of optical properties for biological media can be found in Chapter 5 of [207]). Reconstruction images of μ_a and μ'_s are presented in Fig. 9.7. As in the case of the first two phantoms, including the diffuse phantom, the reconstructions obtained with the SP_3 models are more accurate than those obtained with the SP_1 model.

In contrast to the diffuse phantom (Section 9.5.3), the difference in accuracy between the SP_1 and SP_3 models are only marginal, as the correlation coefficients of the reconstructions obtained

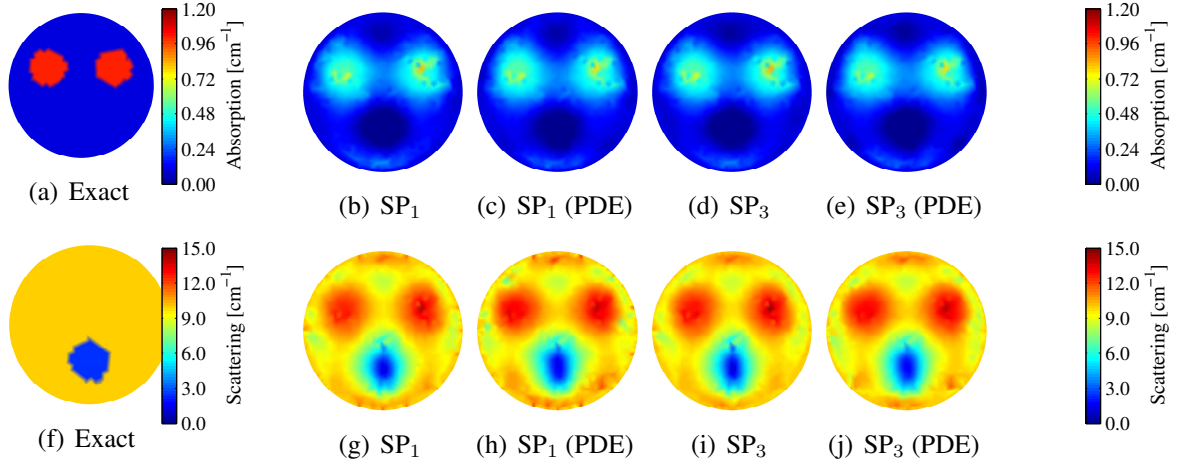


Figure 9.7: (Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).

with the SP_3 model are only 0.02 larger than the correlation coefficients of reconstructions with the SP_1 model. Furthermore, the accuracy of the μ_a and μ'_s reconstruction are similar, with a correlation coefficient between 0.71 and 0.73 (Table 9.5).

Computationally, the PDE-constrained algorithm performs significantly better than the unconstrained algorithm, achieving speedup factors of approximately 12.3 and 15.2 with the SP_1 and SP_3 models, respectively (Fig. 9.8).

Table 9.5: Correlation coefficient between reconstruction and analytical solutions for the third phantom.

	SP_1	SP_1 (PDE)	SP_3	SP_3 (PDE)
$c(\mu_a)$ [%]	0.71	0.72	0.73	0.73
$c(\mu'_s)$ [%]	0.71	0.71	0.73	0.73

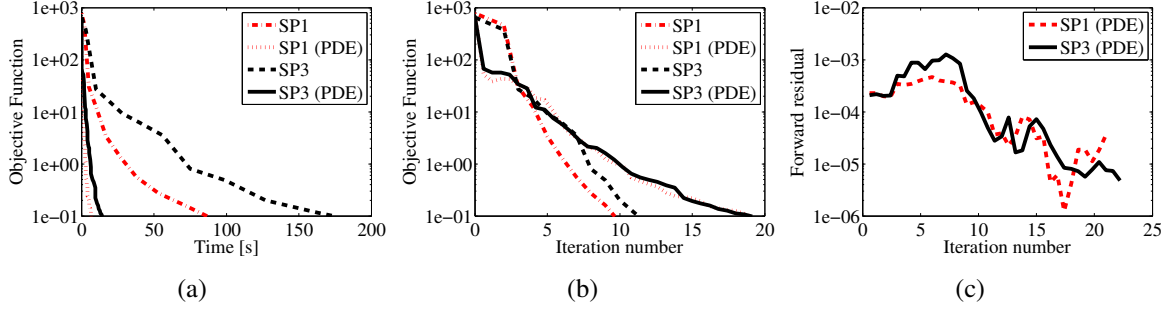


Figure 9.8: Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.

9.5.6 Disk phantom: case 4

The optical properties of this phantom are chosen to be well outside the diffuse regime, with two absorbing inclusions ($\mu_a = 0.2 \text{ cm}^{-1}$) and two regions with low scattering values ($\mu'_s = 6.0 \text{ cm}^{-1}$) in a moderately diffuse background ($\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu'_s = 10.0 \text{ cm}^{-1}$) (Fig. 9.9a,f).

Reconstruction images of the μ'_s are more accurate than reconstructions of the μ_a , as the scattering objects are very accurately resolved while the absorbing objects appear as single continuous objects (Fig. 9.9). The SP_3 model achieves better resolution of the absorbing objects compared to the SP_1 model. In general, the reconstructions with the SP_3 model are more accurate than the

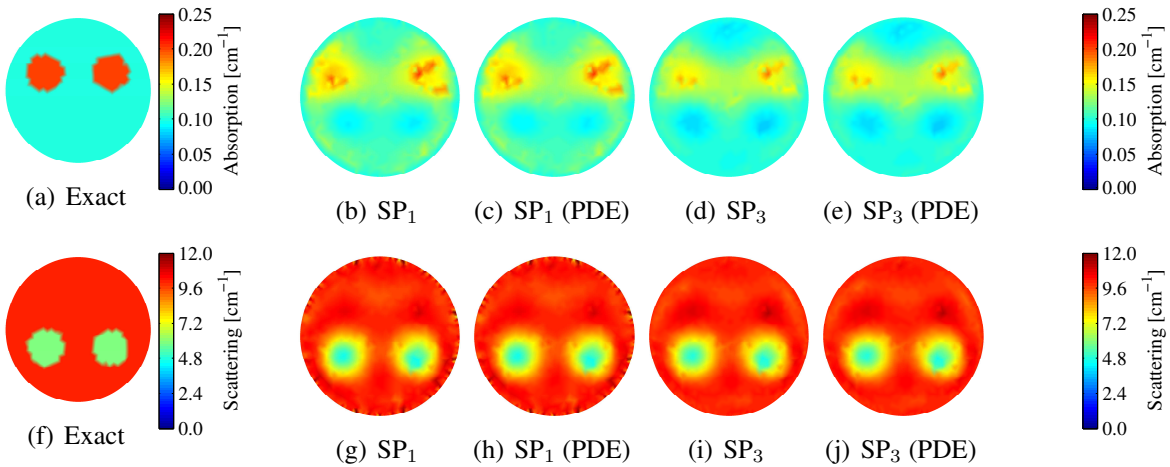


Figure 9.9: (Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).

Table 9.6: Correlation coefficient between reconstruction and analytical solutions for the fourth phantom.

	SP ₁	SP ₁ (PDE)	SP ₃	SP ₃ (PDE)
$c(\mu_a)$ [%]	0.64	0.64	0.67	0.67
$c(\mu'_s)$ [%]	0.77	0.77	0.87	0.87

reconstructions with the SP₁ model (Table 9.6).

Computationally, the PDE-constrained algorithm performs significantly better than the unconstrained algorithm, achieving speedup factors of approximately 9.6 and 9.7 with the SP₁ and SP₃ models, respectively (Fig. 9.10).

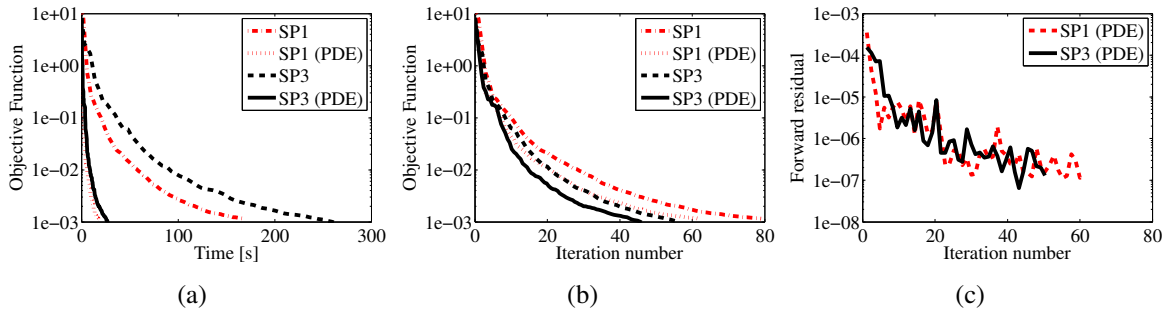


Figure 9.10: Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.

9.5.7 Disk phantom: case 5

In this phantom we test the performance of the algorithms on media with moderately absorbing inclusions ($\mu_a = 0.2 \text{ cm}^{-1}$) and two regions with low scattering values ($\mu'_s = 4.0 \text{ cm}^{-1}$) in a moderately diffuse background ($\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu'_s = 10.0 \text{ cm}^{-1}$) (Fig. 9.11a,f). This phantom is similar to the fourth phantom (Section 9.5.6), with the only difference being the value of the low scattering regions. As in the case of the fourth phantom, the SP₃ model performs better than the SP₁ model and the μ'_s reconstruction is more accurate than the μ_a reconstruction (Fig. 9.11).

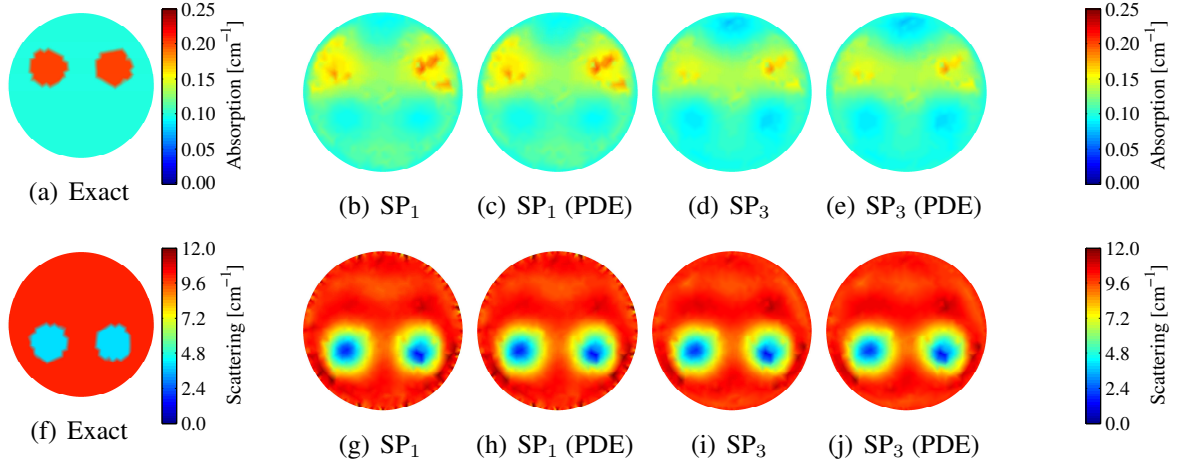


Figure 9.11: (Top) μ_a and (bottom) μ'_s distribution in the two-dimensional phantom. (a,f) Exact distribution, (b,g) SP_1 (unconstrained), (c,h) SP_1 (PDE-constrained), (d,i) SP_3 (unconstrained), and (e,j) SP_3 (PDE-constrained).

In contrast to the fourth phantom, where the SP_3 model performed significantly better than the SP_1 model, the differences in accuracy between the SP_1 and SP_3 models are smaller in this example (Table 9.7). The accuracy of the μ_a reconstruction decreases compared to reconstruction of the phantom in Section 9.5.6, where the scattering inclusions were defined by $\mu'_s = 6.0 \text{ cm}^{-1}$. The accuracy of the μ'_s reconstructions with the SP_3 model remains relatively unchanged, while the accuracy of the μ'_s reconstructions with the SP_1 model improves.

This change is presumed to occur as a result of the significant decrease in the scattering values of the inclusions ($\mu'_s = 6.0 \text{ cm}^{-1}$ to $\mu'_s = 4.0 \text{ cm}^{-1}$). It is expected that approximations to the ERT will perform increasingly poorly as the optical properties of the tissue approach low-scattering media (i.e. void-like).

Computationally, the PDE-constrained algorithm performs significantly better than the uncon-

Table 9.7: Correlation coefficient between reconstruction and analytical solutions for the fifth phantom.

	SP_1	SP_1 (PDE)	SP_3	SP_3 (PDE)
$c(\mu_a)$ [%]	0.61	0.61	0.62	0.62
$c(\mu'_s)$ [%]	0.82	0.82	0.86	0.86

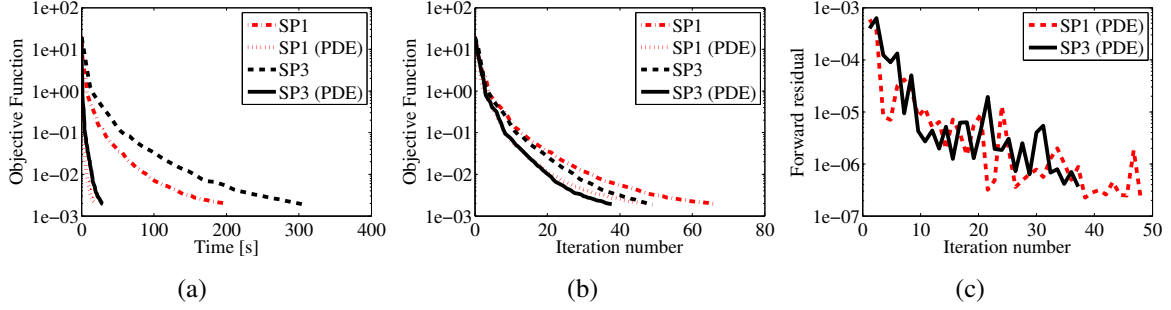


Figure 9.12: Performance profile of reconstruction algorithms. Objective function value over (a) time and (b) at the conclusion of each inverse iteration. (c) Evolution of the forward model residual associated with the PDE-constrained algorithm across each inverse iteration.

strained algorithm, achieving speedup factors of approximately 11.1 and 10.5 with the SP_1 and SP_3 models, respectively (Fig. 9.12).

9.5.8 Finger phantom

In this section we consider a three-dimensional numerical phantom that is based on the surface geometry of a human finger. The source and detector configuration of this phantom is modeled after a trans-illumination imaging setup, where the finger is illuminated on the posterior (dorsal) side and escaping photons are measured on the anterior (palmar) side. In total, 11 distinct point sources (Fig. 9.13a) and 155 detector points (Fig. 9.13b) are defined for this phantom.

Two cases of the finger phantom are considered. In both cases we defined one inclusion with elevated absorption and one inclusion with lower scattering than the background medium (Fig. 9.14). The section of the finger under consideration is approximately 4.0 cm in length (y -axis) and has width and height of approximately 2.0 cm. The inclusions are spheres with a 1.5 cm diameter. The reconstruction process is allowed to continue until the inverse error decreases by a factor of 0.0001 (or to 0.01% of the original value).

Optical properties are chosen to closely resemble those typically found in and around the prox-

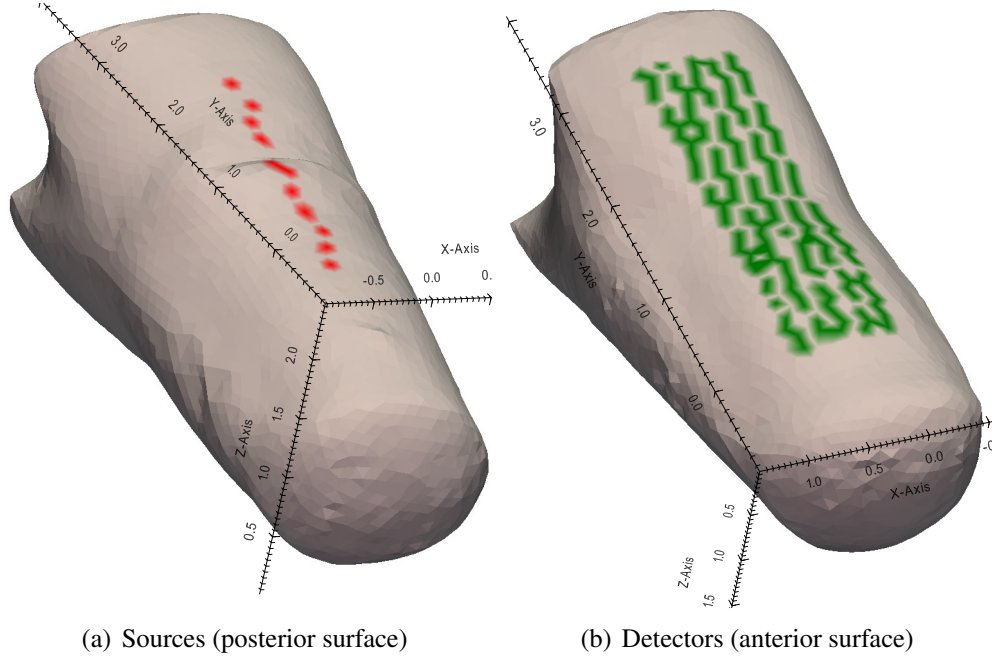


Figure 9.13: (a) Position of the 11 sources and (b) location of the 155 detectors on the posterior and anterior surface of the finger phantom, respectively.

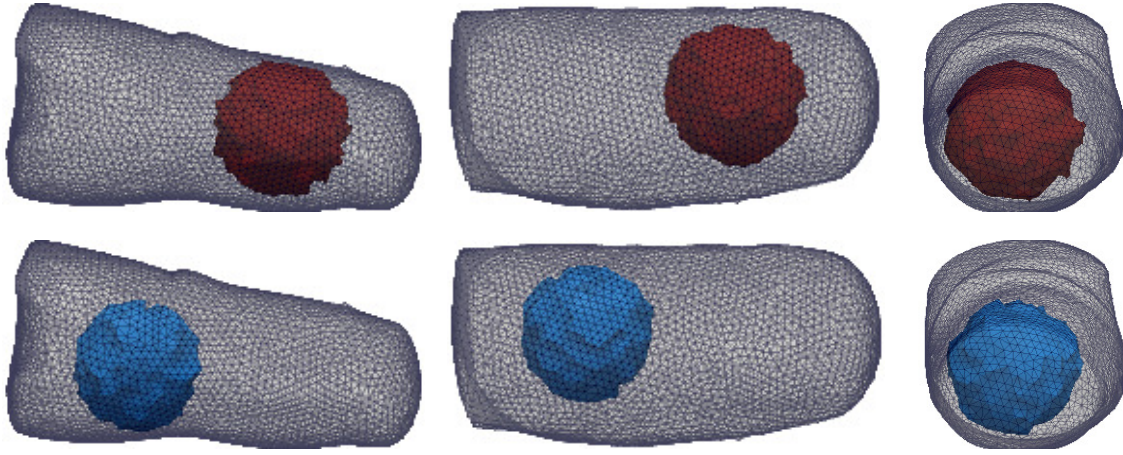


Figure 9.14: Definition of finger-like phantoms; both phantoms have one absorbing (red object, top row) and one scattering (blue object, bottom row) inclusion in a homogeneous background. The optical properties of the first phantom are as follows; the background properties are $\mu_a = 0.01 \text{ cm}^{-1}$ and $\mu'_s = 12.5 \text{ cm}^{-1}$ ($g = 0.95$), the absorbing inclusion is $\mu_a = 0.05 \text{ cm}^{-1}$, and the scattering region is $\mu'_s = 12.5 \text{ cm}^{-1}$. The second finger phantom is similar, except for differences in the optical properties of the background ($\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$, $g = 0.95$), the absorbing inclusion ($\mu_a = 0.25 \text{ cm}^{-1}$), and the scattering region ($\mu_s = 7.5 \text{ cm}^{-1}$).

Table 9.8: Correlation coefficient between reconstruction images and analytical solution for the first finger phantom (one absorbing object with $\mu_a = 0.05 \text{ cm}^{-1}$ and one scattering object with $\mu'_s = 10.0 \text{ cm}^{-1}$).

Measure	SP ₁	SP ₁ (PDE)	SP ₃	SP ₃ (PDE)
$c(\mu_a)$ [%]	0.54	0.54	0.67	0.66
$c(\mu'_s)$ [%]	0.60	0.60	0.71	0.71

Table 9.9: Correlation coefficient between reconstruction images and analytical solution for the second finger phantom (one absorbing object with $\mu_a = 0.25 \text{ cm}^{-1}$ and one scattering object with $\mu'_s = 7.5 \text{ cm}^{-1}$).

Measure	SP ₁	SP ₁ (PDE)	SP ₃	SP ₃ (PDE)
$c(\mu_a)$ [%]	0.65	0.65	0.72	0.72
$c(\mu'_s)$ [%]	0.60	0.60	0.67	0.67

imal interphalangeal (PIP) joint of the human hand [9]. The inclusions are chosen to resemble regions of high absorption and low scattering that can be found in and around the PIP joint cavity.

In the first case, the background media is defined with $\mu_a = 0.01 \text{ cm}^{-1}$ and $\mu'_s = 12.5 \text{ cm}^{-1}$ ($g = 0.95$); the absorption inclusion has $\mu_a = 0.05 \text{ cm}^{-1}$ and the scattering inclusion has $\mu'_s = 10.0 \text{ cm}^{-1}$ (Fig. 9.14). The second finger phantom is similar, with only differences in the value of the optical properties. The background media is defined with $\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$ ($g = 0.95$); the absorption inclusion has $\mu_a = 0.25 \text{ cm}^{-1}$ and the scattering inclusion has $\mu_s = 7.5 \text{ cm}^{-1}$. The reconstruction of these optical properties allows us to determine how accurately each algorithm can determine the underlying optical properties and their individual computational efficiency. These simulations closely resemble clinical applications.

The accuracy of reconstructions obtained with the SP₁ and SP₃ models are summarized in Tables 9.8 and 9.9, respectively. In both cases, the reconstructions obtained with the SP₃ model are more accurate than the reconstructions obtained with the SP₁ model. The differences between the models are more significant for the first finger phantom.

An example of the absorption and scattering coefficient maps obtained with the SP₃ model using the PDE-constrained algorithm is presented in Fig. 9.15. In general, the algorithm is able to accurately locate the inhomogeneities and can accurately identify the value of the absorption and

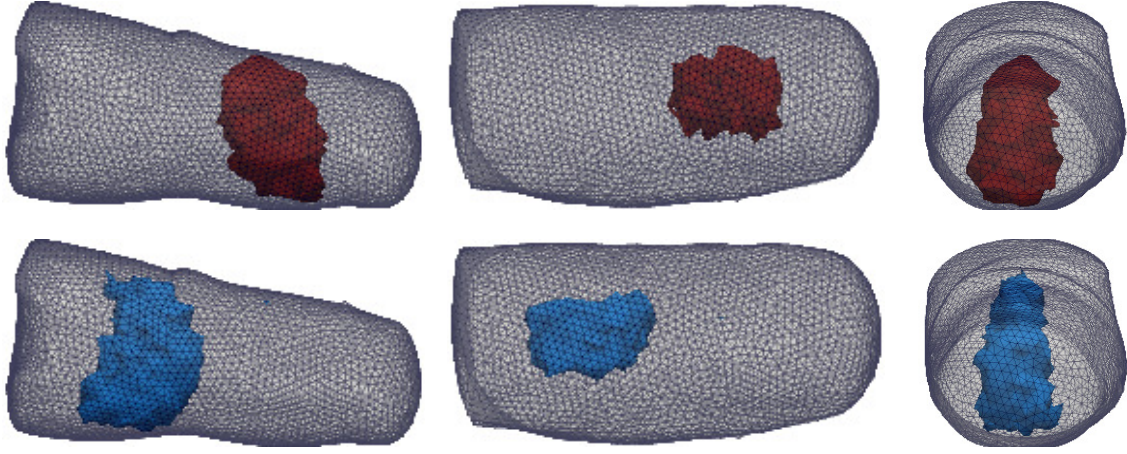


Figure 9.15: Three-dimensional reconstruction images corresponding to the second finger phantom. Depicted are the absorption inclusion (red object) and scattering void (blue object) in otherwise homogeneous media. The true optical properties of the absorbing inclusion is $\mu_a = 0.25 \text{ cm}^{-1}$, the scattering region is $\mu_s = 7.5 \text{ cm}^{-1}$, and the background are $\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$. The reconstruction algorithm is able to accurately locate the inhomogeneities and estimate their optical properties. The phantom is defined in Fig. 9.14.

scattering coefficients in the same general location (Fig. 9.16).

The computational performance of the reconstruction process of both numerical phantoms are summarized by Fig. 9.17, which plots the value of the objective function over time. The PDE-constrained algorithms significantly outperform the unconstrained algorithms in the reconstruction of both phantoms. However, the computational efficiencies gained by the PDE-constrained algorithm is more significant in solving for the first phantom. In that case, computational speedup factors of up to 5 are obtained.

Another computational aspect of interest is the system memory (RAM) requirements of each algorithm as these two finger phantoms closely resemble clinical applications. Reconstructions with the SP_1 model requires up to 177.5 MB, while up to 212.7 MB are used by the SP_3 algorithm. These computational requirements are not significant compared to the RAM found in typical desktop computers (12 GB in the desktop used for these studies), which is important as it is proof that our algorithm is efficient in memory management. This efficiency is expected to translate to general DOT problems and is not specific to the simulations presented in this section. For example, we expect similar efficiencies when considering clinical data.

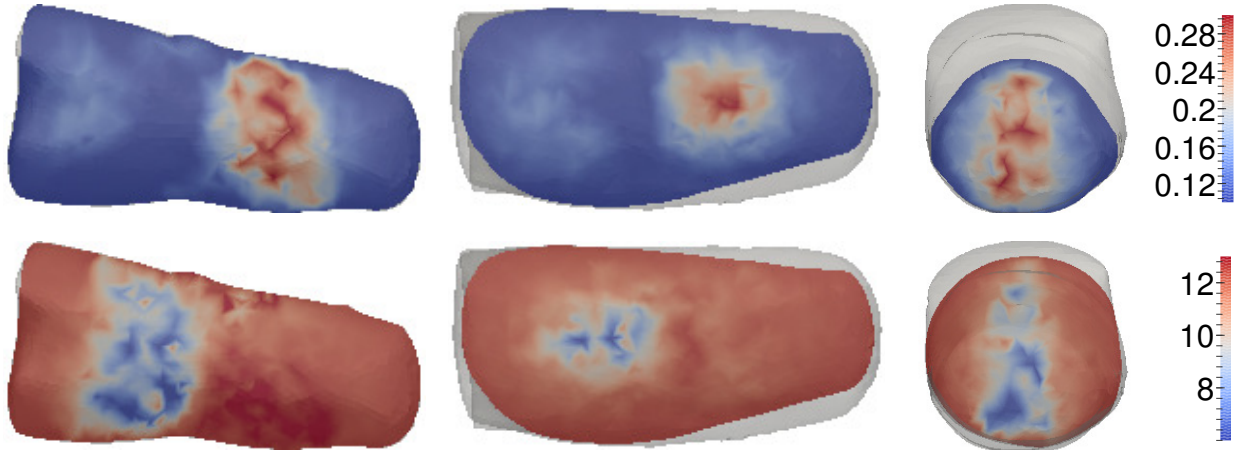


Figure 9.16: Cross-sectional slices inside the finger phantom of the reconstruction images of the absorption (top row) and scattering (bottom row) coefficients obtained with the PDE-constrained SP_3 models for the second finger phantom. The true optical values of this phantom are as follows; background media is defined with $\mu_a = 0.1 \text{ cm}^{-1}$ and $\mu_s = 12.5 \text{ cm}^{-1}$ ($g = 0.95$), the absorption inclusion has $\mu_a = 0.25 \text{ cm}^{-1}$, and the scattering inclusion has $\mu_s = 7.5 \text{ cm}^{-1}$.

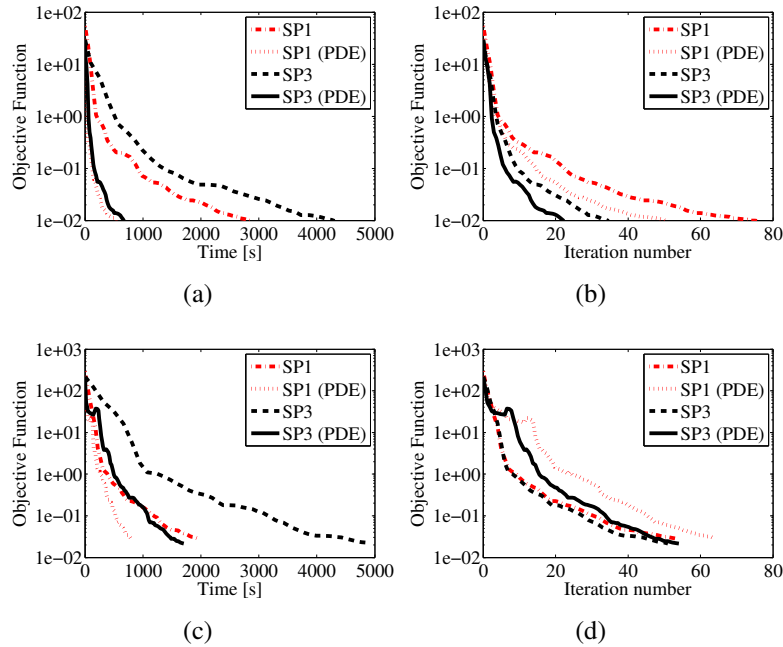


Figure 9.17: Performance profile of reconstruction algorithms. Objective function value over (a,c) time and (b,d) at the conclusion of each inverse iteration for the (top row) first and (bottom row) second finger phantom.

9.6 Discussion

Two image-reconstruction algorithms for use in DOT imaging that are based on solving the finite-volume frequency-domain SP_N model (FV-FD- SP_N) are presented in this chapter. Both optimization methods use the quasi-Newton limited-memory BFGS algorithm to directly compute the Newton direction (Δx), which is then used to directly update the inverse variable in an iterative manner ($\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \Delta \mathbf{x}^k$). Line search based on cubic interpolation is used to find the update step size (α) that achieves sufficient decrease in the merit (or objective) function. One approach is based on unconstrained optimization, where the goal of the algorithm is to repeatedly solve the forward model to iteratively update the inverse variable. The second approach is based on PDE-constrained optimization theory, where the inverse and forward variables are both solved simultaneously and iteratively updated.

The algorithms are validated using numerical phantoms of two- and three-dimensional geometries. Simulated measurement data is created using the SP_3 model on dense numerical grids and corrupted with 1% Gaussian white noise. The reconstruction process is performed on less dense numerical grids using the SP_1 and SP_3 models. The accuracy and computational efficiency of each type of optimization algorithm and SP_N model is analyzed. Optical parameters within the diffuse regime and outside the diffuse regime are considered to understand how the algorithms may perform in clinical data applications, where optical parameters can lie within large ranges.

Overall, the absorption and scattering coefficient maps are more accurately reconstructed with the SP_3 model than with the SP_1 model. The PDE-constrained algorithm is typically more than one order of magnitude faster than the unconstrained algorithm. This speedup factor is obtained without loss of accuracy by the PDE-constrained algorithm. The speedup factor is lower when three-dimensional geometries are considered. We find that the PDE-constrained algorithm performs well when the adjoint variable and Newton direction for the forward variable are both solved with the GMRES algorithm with outer and inner loop tolerances of 0.001.

We note that the difference in computation time between the algorithms that use the SP_1 and SP_3 model is minimal when using the PDE-constrained approach. While a typical inverse iteration

with the SP_1 model is completed in about half the time it takes for an SP_3 iteration, the reconstruction process with the SP_3 model typically requires fewer inverse iterations than the process that uses the SP_1 model; these convergence properties result in reducing the overall difference in reconstruction time between the SP_1 and SP_3 models.

The focus of Chapter 10 is the application of the PDE-constrained optimization algorithm to clinical data. We focus on investigating the utility of the SP_3 algorithm when considering clinical data. Modifications necessary to incorporate clinical data into the algorithm are presented. Finally, results from the application of computer added diagnosis (CAD) techniques to the reconstruction data are presented.

Application Of The PDE-Constrained SP_N Algorithm To Clinical Data

The focus of the work reported in this chapter is to combine the image analysis and DOT reconstruction tools presented thus far to study our ability to diagnose RA using the SP_N light propagation model. First, we use the SP_N reconstruction algorithm to recover the absorption and scattering parameters from the finger joint clinical data introduced in Chapter 5, which consists of frequency domain DOT images of 219 peripheral interphalangeal (PIP) joints. Then, we use the computer-aided diagnosis (CAD) algorithm presented in Chapters 6 and 7 to investigate our ability to diagnose RA using the reconstruction images obtained with the SP_N model.

We have at our disposal clinical data captured at 600, 300, and 0 MHz, as well as reconstruction algorithms with an unconstrained or a constrained optimization approach. We also have a large suite of classification techniques available, as presented in Chapters 6 and 7. Instead of investigating all possible scenarios, we use results from Chapters 5, 6, 7, 8, and 9 to make the following choices.

We choose to use clinical data captured with modulation frequency of 600 MHz. The choice of modulation frequency is made based on evidence shown in Chapter 5 that indicates that 600 MHz data leads to higher sensitivities and specificities than data at 300 and 0 MHz.

The PDE-constrained reconstruction algorithm is used instead of the unconstrained algorithm

for the two reasons that are discussed in detail in Chapter 9. Primarily, the PDE-constrained reconstruction algorithm has superior computational performance compared to the unconstrained algorithm, achieving speed-up factors between 5 and 20 times without loss in accuracy.

Lastly, we choose to use the support vector machine (SVM) algorithm in our feature extraction and optimal feature selection algorithm, as presented in Chapter 7. This choice is based purely on the results presented in Chapter 7, where it is shown that classification based on SVM with a polynomial kernel¹ yields the highest sensitivities and specificities in the diagnosis of RA.

As a final analysis step, we will report on the ability to diagnose RA from reconstructions obtained with the SP_N model and compare them to the ability to diagnose RA from reconstructions obtained using the equations of radiative transfer (ERT), a more accurate light propagation model.

We begin this chapter with Section 10.1, where we introduce modifications to the PDE-constrained reconstruction algorithm presented in Chapter 9 that are necessary for the application to clinical data. Then, in Section 10.2, we provide a brief review of the clinical data that forms a basis to the analysis presented in this chapter. Results from performing image reconstructions on the clinical data with the PDE-constrained reconstruction algorithm are presented in Section 10.3. Results from the application of the CAD algorithm to the SP_N based reconstruction images are presented and discussed in Section 10.4. Furthermore, these results are directly compared to classification results obtained with ERT-based images. This chapter concludes with Section 10.5, a summary of the results presented in Chapter 10 and a discussion on the impact these results may have on the field of diffuse optical tomography.

10.1 Objective function for use with clinical data

Due to practical considerations, when working with experimental data, the objective function of the inverse problem must be modified from the version presented in Chapter 9. First, recall the general form of the objective function that is used to solve the unconstrained optimization problem, given

¹The optimal kernel is a third-order polynomial.

by

$$f(\mu, u) = \frac{1}{2} (M - P)^T \overline{(M - P)}. \quad (10.1)$$

This function must be modified because, in general, the absolute strength of the boundary source ($\text{W cm}^{-2} \text{ sr}^{-1}$) (i.e. laser) is not precisely known. As a result, it is impossible to directly compare the predicted measurement values (P) with the true measurement data (M). The inability to directly compare these vectors is due to a severe mismatch in scale that can occur when the source strength is not precisely known. As a result, the true measurement data is often normalized by the average overall measured intensity.

In the remainder of this section we specify the necessary changes to the standard DOT objective function and how those changes impact the rest of the reconstruction algorithm. First, the normalized measurement data (\tilde{M}) is given by

$$\tilde{M}_{ij} = (\overline{M})^{-1} M_{ij}. \quad (10.2)$$

The indices $i \in S$ and $j \in D$ denote all possible sources and detectors, respectively. Here, \overline{M} represent the average measured intensity, which is a complex valued number when considering frequency domain DOT, and is given by

$$\overline{M} = \frac{1}{SD} \sum_{i=1}^S \sum_{j=1}^D (M)_{ij}. \quad (10.3)$$

Then, mirroring the normalization of M , the predicted detector measurements P must be normalized in a similar manner. The normalized predicted partial current can be computed as follows,

$$\tilde{P} = (\overline{P})^{-1} P. \quad (10.4)$$

Here, \overline{P} is the average predicted partial current across all source-detector pairs and is defined as

$$\overline{P} = \frac{1}{SD} \sum_{i=1}^S \sum_{j=1}^D (Qu)_{ij} \quad (10.5)$$

After incorporating these changes, the augmented objective function can be stated as follows,

$$f(\mu, u) = \frac{1}{2} (\tilde{M} - \tilde{P})^T \overline{(\tilde{M} - \tilde{P})}. \quad (10.6)$$

This change necessitates several specific considerations. In particular, the right hand side of the adjoint equation is more complicated because \overline{P} itself is a function of u , the variable over which differentiation occurs. The adjoint equation for the constrained and unconstrained approach is discussed in detail in Section 9.2.1 and given by (9.21). A numerical recipe for using the solution to the adjoint equation in the PDE-constrained algorithm is described in Section 9.3.1 and given by (9.49). Then, given the changes introduced as a result of normalizing measurement data, the correct adjoint equation is now (a detailed derivation is provided in Appendix B.8),

$$A^T \lambda = Q^T \left(\frac{\tilde{P}^T (\tilde{M} - \tilde{P})^* \frac{1}{SD} \mathbf{1} - (\tilde{M} - \tilde{P})^*}{\overline{P}} \right), \quad (10.7)$$

Normalization of each element of the error (or difference) vector $(\tilde{M} - \tilde{P})_{ij}$ by the corresponding norm of the measurement element (i.e. $|\tilde{M}|_{ij}$) often leads to improved convergence (where $i \in S$ and $j \in D$ denote the source-detector pair). In this case, the objective function is

redefined as,

$$f(\mu, u) = \frac{1}{2} \frac{\left(\tilde{M} - \frac{Qu}{\bar{P}} \right)^T \left(\tilde{M} - \frac{Qu}{\bar{P}} \right)}{\tilde{M}^T \tilde{M}}. \quad (10.8)$$

The only change to Eq. 10.7 is the use of the appropriate normalized terms when computing the difference between the predicted measurements and the true measurements, $\left(\tilde{M} - \tilde{P} \right)$.

10.2 Clinical data

Details of the clinical data can be found in Chapter 5. To summarize, the clinical data consists of 219 PIP joints that were imaged using a sagittal trans-illumination frequency-domain DOT scanner. PIP joints II-IV were imaged on the dominant hand of 33 subjects with RA and on both hands of 20 healthy control subjects, resulting in 99 joints from subjects with RA and 120 joints of subjects without RA.² Scanning was performed at 600, 300, and 0 MHz. The source laser illuminated the joint on the posterior (dorsal) side and escaping photons were measured on the anterior (palmar) side using an intensified CCD-based detection system. In total, 11 distinct point sources (Fig. 10.1a) and over 100 detector points (Fig. 10.1b) were defined for each joint.

An example of transillumination from a single surface source, as captured by the ICCD-based detector system, on the posterior (or palmar) surface of the finger is presented in Fig. 10.2. The value of the detector mesh points (Fig. 10.1b) are obtained by mapping those pixels onto the ICCD image and then extracting the measured data at the corresponding pixels.

²A total of 36 subjects with RA were scanned, however, data from three subjects was discarded due to errors that occurred during the DOT imaging procedure.

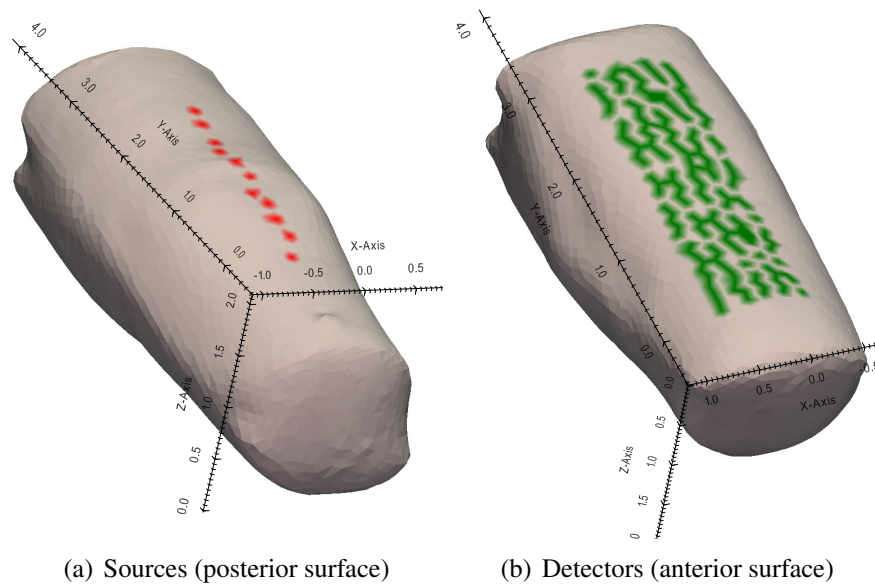


Figure 10.1: An illustration of a typical finger section. The PIP joint is located approximately in the middle of the object, whose length is approximately 4 cm. (a) The position of the 11 unique sources are indicated on the posterior surface of the finger. (b) The position of the detectors (i.e. mesh nodes that map directly to a pixel on the CCD camera-based detection system) are shown on the anterior surface of the finger section.

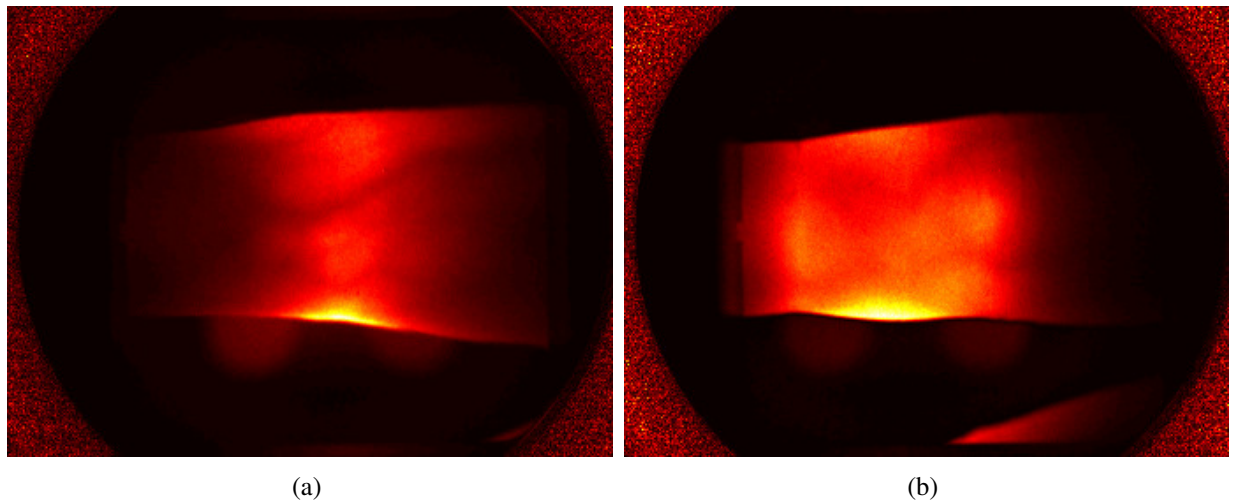


Figure 10.2: Transillumination captured by the ICCD detector unit on the posterior (palmar) surface of a PIP joint belonging to (a) a subject with RA and (b) a healthy control.

10.3 Reconstruction results

Reconstruction of the absorption and scattering coefficients of each imaged joint is performed using the PDE-constrained SP_3 algorithm. Qualitatively, the images appear as expected, with regions of low absorption and scattering within the area where we expect the PIP joint cavity.

Each reconstruction requires between 150 to 200 MB of RAM and is completed in less than 15 minutes. The objective function typically decreases to between 1.0% and 10.0% of the original objective function value. We note that the 10,000 fold reduction of the inverse error from its original value that we achieve in the numerical simulations (Section 9.5.8) is not, in general, obtained in the reconstructions of clinical data. The inability to obtain more substantial reductions in the objective function can be attributed to the loss of information that arises from normalizing the measurement data by the average overall measured intensity.

This is a necessary step, however, as the reconstruction results without normalization can be significantly more inaccurate and unreliable when using clinical data. Indeed, attempting to reconstruct clinical data without performing the aforementioned normalization often leads to early termination of the reconstruction process (i.e. the algorithm cannot progress beyond the initial guess) or if left unchecked, the value of the objective function can increase dramatically (this only occurs if we do not add specific constraints on allowable values of the objective function, such as requiring it to constantly decrease at each inverse iteration).

The reconstruction images of absorption and scattering within and around the imaged PIP joint show distinct differences between subjects with RA and the control group. The μ_a and μ'_s images of 40 joints are presented in Figs. 10.3 and 10.4. In these images, it is clear that compared to joints of healthy subjects, subjects with RA have elevated regions of absorption and scattering around the PIP joint.

In these images, a region of low scattering and absorption is identifiable around the location where we expect the PIP joint. This region may correspond to the synovial cavity of the joint, which we expect has lower absorption and scattering compared to the surrounding tissues (bone, muscle, ligament, tendon). This region is identifiable in all joints of healthy subjects and some

joints of subjects with RA. However, these same regions exhibit elevated absorption and scattering (compared to the background) in joints of subjects with RA. This may suggest that the synovial cavity of these joints has experienced significant changes in physiology that results in an elevated concentration of absorbers and scatterers; such changes can be expected due to the onset of symptoms associated with RA.

The reconstruction parameters used to obtain these results are summarized by Table 10.1. The absorption and scattering coefficients were rescaled by transforming the absorption range of $[0.0, 0.4] \text{ cm}^{-1}$ and the scattering range of $[0.0, 400.0] \text{ cm}^{-1}$ to $[0.0, 1.0] \text{ cm}^{-1}$. We note that this rescaling does not imply that we restrict the acceptable solutions to lie within this range. Indeed, the acceptable range of solutions itself is not restricted.

Table 10.1: Summary of values assigned to reconstruction variables.

Parameter Name	Value
Background μ_a	0.3 cm^{-1}
Background μ_s	200.0 cm^{-1}
Anisotropy factor g	0.95
Refractive index of medium n	1.4
Refractive index of air n	1.0
Modulation frequency ω	600.0 MHz
Speed of light in medium c	299.87×10^8
GMRES forward model tolerance (τ)	1.0×10^{-12}
GMRES adjoint model tolerance (τ)	1.0×10^{-4}
GMRES inexact forward model tolerance (τ)	1.0×10^{-4}
GMRES maximum number of iterations	500
GMRES restart iterations (mr)	50
Boundary regularization	No
Inverse tolerance	0.01
Minimum decay rate	1.0×10^{-8}
Boundary source power	$1.0 \text{ W cm}^{-2} \text{ sr}^{-1}$
Discrete ordinates	S_{12} (168)
Absorption rescaling range	$[0.0, 0.4] \text{ cm}^{-1}$
Scattering rescaling range	$[0.0, 400.0] \text{ cm}^{-1}$

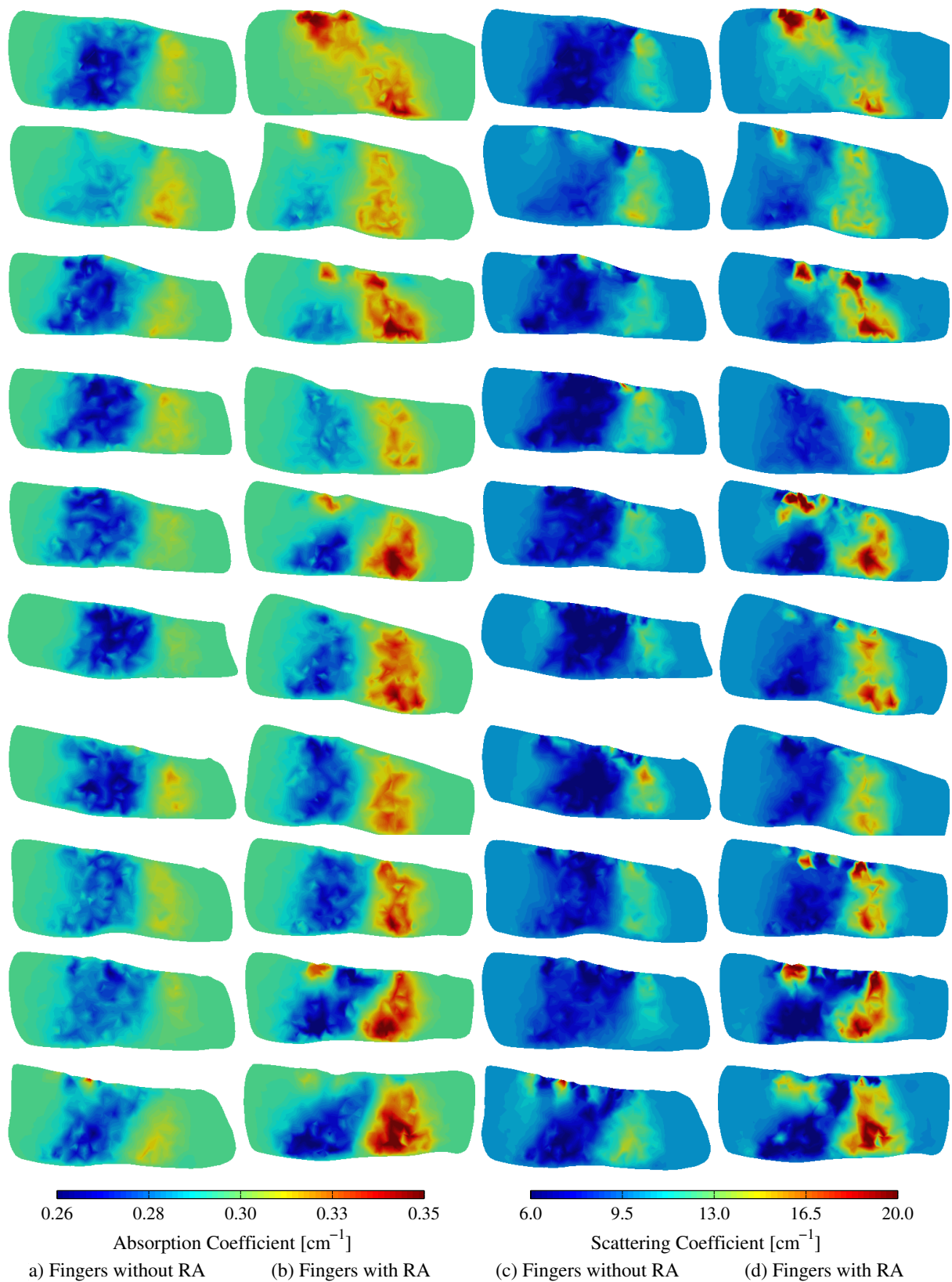


Figure 10.3: (a-b) Absorption and (c-d) scattering cross-sections of distinct joints from subjects with and without RA.

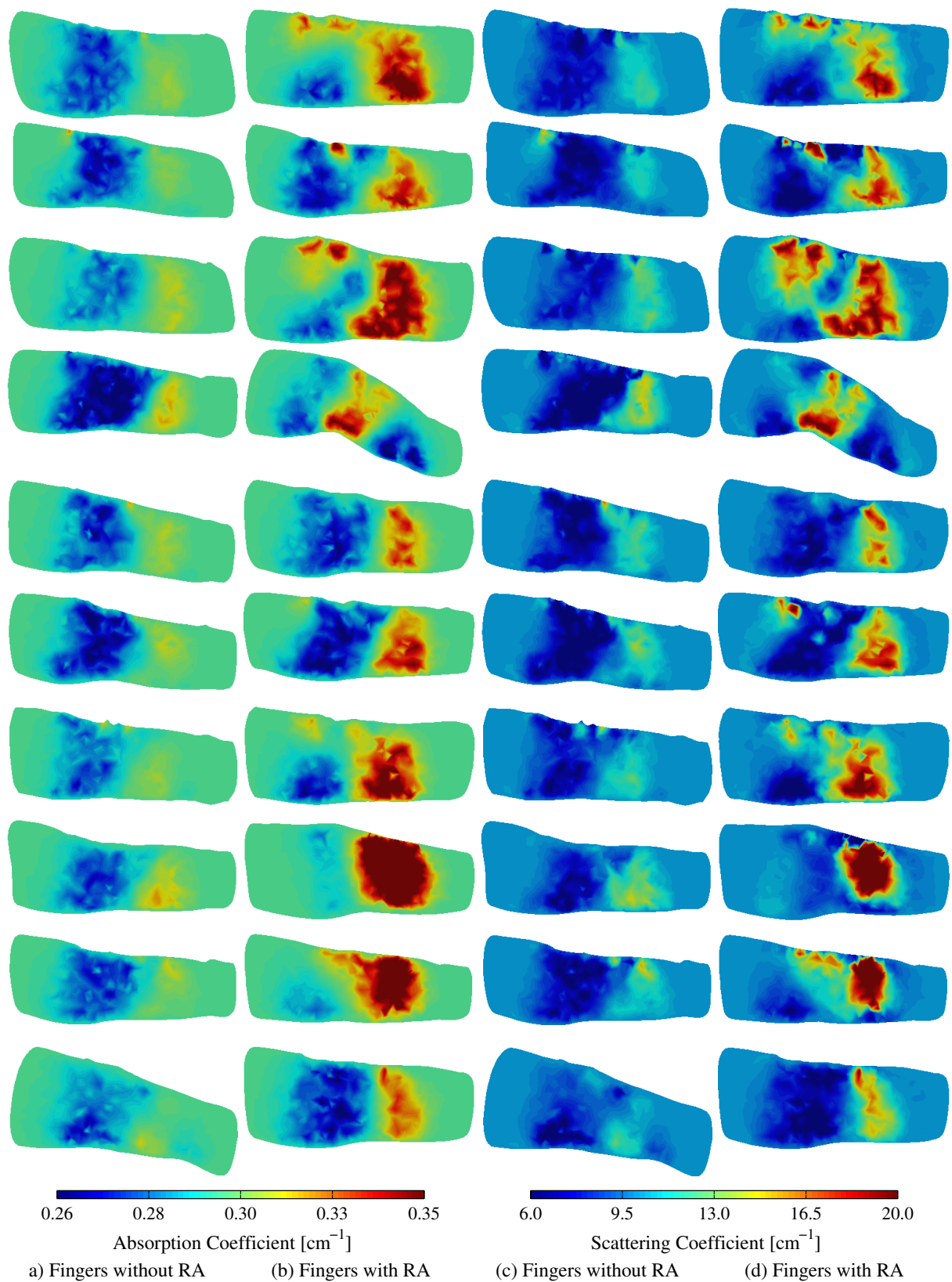


Figure 10.4: (a-b) Absorption and (c-d) scattering cross-sections of distinct joints from subjects with and without RA.

10.4 Analysis of SP_N reconstruction data with CAD

The next step in our analysis is to use reconstruction results obtained with the SP_N model to diagnose RA. We use the k -fold cross-validation technique to gauge the ability to diagnose RA with this new data. Here, approximately 2/3 of the data is used to train and 1/3 is used to test. The training set consists of 22 subsets with RA (or 66 PIP joints) and 14 healthy individuals (or 84 PIP joints), together consisting of 150 distinct joints. The testing set includes 11 subjects with RA (or 33 PIP joints) and six healthy patients (or 36 PIP joints). The allocation of subjects into the training of the testing group is done randomly to help minimize any potential bias.

The cross-validation process used in Chapters 6 and 7 is modified in two small, yet important ways. First, the rule for selecting the top 30 features for use in the optimization algorithm is changed from the classical Youden index (i.e. $Y = Se + Sp - 1$) to the augmented Youden index introduced in Section 7.2.6, given by

$$Y^*(Se, Sp) = Se + Sp + \alpha L_{Se} + \beta L_{Sp} - \delta d - 1. \quad (10.9)$$

Here, the Youden index is a function of sensitivity (Se) and specificity (Sp), and the lower bounds of the confidence intervals of Se and Sp , given by L_{Se} and L_{Sp} , respectively.

The second modification is the use of the k -fold method to enhance the cross-validation process. With $k = 2$, we segment the entire data set into two sets, a training and a testing set. The training set is subjected to the data-mining process presented in Chapters 6 and 7, which yields a set of optimal feature vectors and the corresponding choice of classification algorithm for the training data. The testing set is not used until the optimal classifier (features and classification algorithm) is chosen. The resulting classifier is used to subsequently evaluate the ability to diagnose RA with the testing data. The process is summarized by the flowchart in Fig. 10.5.

Thus, the only difference between this analysis and the analysis in Chapters 6 and 7 is the additional layer of cross-validation that arises from keeping a set of data away from the data mining

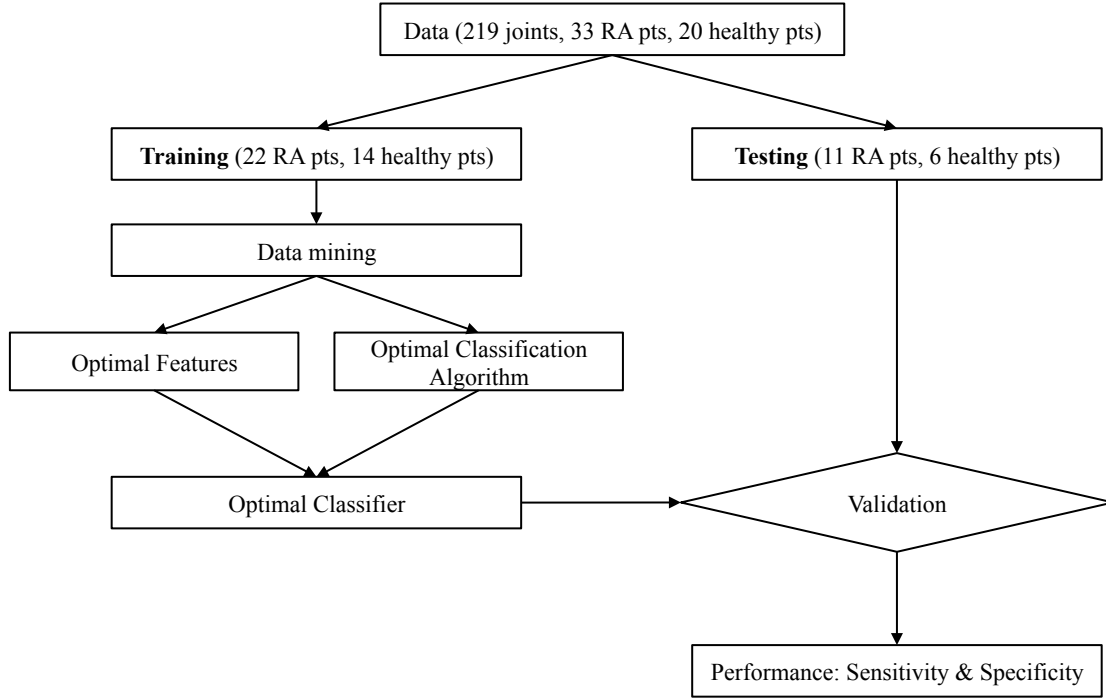


Figure 10.5: Flow chart of k -fold cross-validation.

process and then subsequently applying the results from the optimization algorithm on the testing set. This so-called k -fold process provides additional safety layers to ensure our classification results are as unbiased as possible and thus more reliable.

As a reminder, the data mining process presented in Chapter 7 is summarized by the flowchart in Fig. 10.6. This process is performed on reconstruction images obtained with the SP_1 , SP_3 , and ERT models.³

For each data set (i.e. SP_1 , SP_3 , or ERT based reconstructions) we report the sensitivity and the specificity obtained when the data mining results are used to classify the testing data. As in Chapter 7, we report the 95% confidence intervals (CI) associated with the sensitivity and the specificity. As a reminder, the CI informs us of the lower and upper bounds of the sensitivity and the specificity within which we are confident the true performance of the classifier is located (with up to 95% confidence). The CI is computed to account for the effective sample size of our data,

³As a reminder, the ERT-based reconstructions are the original data set and are first discussed in Chapters 5, 6, and 7.

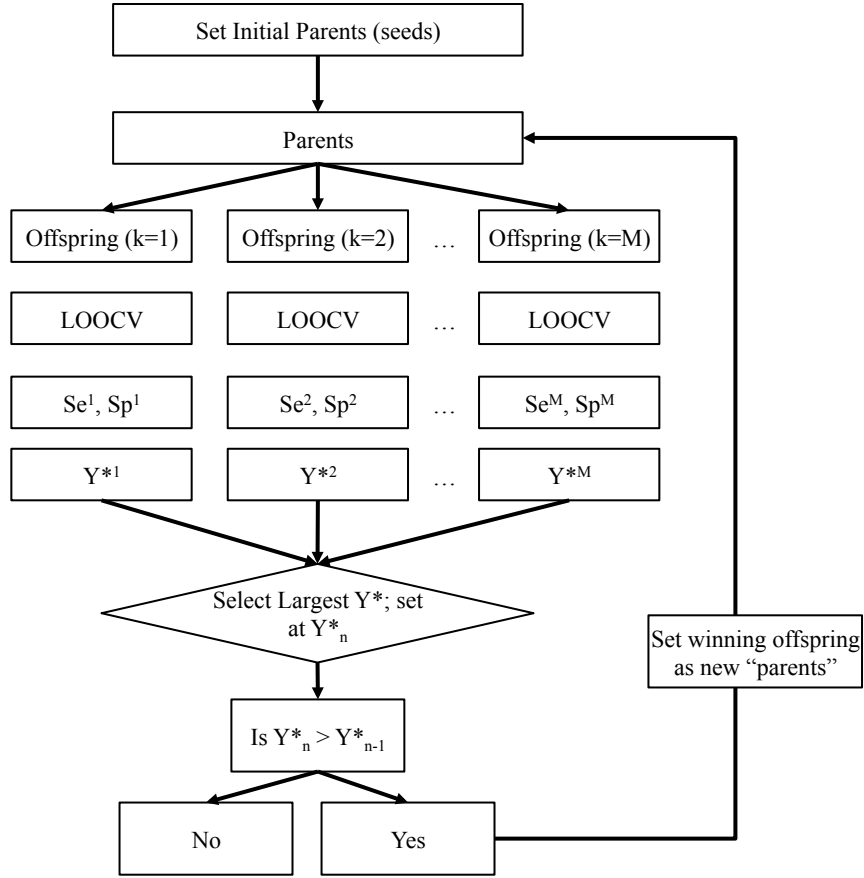


Figure 10.6: Flow chart of the data mining process that selects optimal features and the best classification algorithm for diagnosing RA from DOT images.

which removes bias that may be introduced into our statistical parameters through correlation between joints of the same subject. See Section 6.4.2 for details on the computation of the confidence intervals.

Classification results are summarized in Tables 10.2 and 10.3. The sensitivity and specificity (and their respective confidence intervals) achieved when using the reconstruction images generated with the SP_1 , SP_3 , and ERT models are summarized in Table 10.2. The number of true positives, false negatives, true negatives, and false positives are also summarized in Table 10.2.

We consider three criteria for comparing the performance of the SP_1 , SP_3 , and ERT models. The three categories are computational efficiency, feature extraction optimality, and image classification performance.

Table 10.2: Classification results for SP_N and ERT based reconstructions with the k -fold method.

Model	TP	FN	TN	FP	Se [% (95% CI)]	Sp [% (95% CI)]	Youden	Number of Features
SP1	22	11	34	8	66.7 (46.6,100)	81.1 (64.8,100)	0.48	8
SP3	29	4	39	3	87.9 (78.1,100)	92.9 (84.6,100)	0.81	3
ERT	30	3	41	1	90.9 (83.1,100.0)	97.6 (85.1,100.0)	0.88	5

The first category in which we compare the models is computational efficiency, which consists of the total reconstruction time and the total system resources needed. The reconstruction times with the SP_1 and SP_3 models are similar, typically requiring less than 15 minutes to complete the reconstruction process. In contrast, the reconstruction time with the ERT model always exceeds 180 minutes.⁴ Additionally, reconstructions with the ERT always required over 6 GB of RAM, while the SP_1 and SP_3 models always required less than 200 MB of RAM. In the computational efficiency category the SP_N models outperform the ERT model.

The second category is feature extraction optimality, where we compare the number of “optimal” features selected during the training process. In general, we prefer optimal feature vectors with low-dimensionality as this helps reduce the probability of over-fitting the data. Over-fitting the data can result in classification results that do not generalize well and therefore may be an unreliable predictor of future performance. In this work, the number of optimal features is eight for the SP_1 model, three for the SP_3 model, and five for the ERT model. As we are generally interested in the fewest possible features to avoid over fitting problems, it is clear the SP_3 model is superior in this category to the ERT and SP_1 models.

The third category to compare is the classification performance of each of the three models. Here, we are primarily concerned with the sensitivity and the specificity that is computed by processing the data set reserved for testing with the classifier that results from the training phase. In addition to seeking values of the sensitivity and the specificity as close to 100.0% as possible, we

⁴Indeed, this reconstruction time is actually orders of magnitude faster than traditional ERT-based reconstruction algorithms that do not solve the optimization problem using a PDE-constrained approach. Furthermore, these ERT-based reconstruction were obtained with a low order discrete ordinates set (S_6), which is significantly lower than the recommended set of ordinate to accurately capture transport behavior ($> S_{12}$).

are also interesting in comparing the 95% confidence interval for each parameter. The CI is important because it informs us of the range within which we expect the true values of the sensitivity and the specificity to exist.

The reconstruction images computed with the SP_3 model allow higher sensitivity and specificity values than images obtained with the SP_1 model. The SP_1 model yields sensitivity of 66.7 (46.6, 100.0)% and specificity of 81.1 (64.8, 100.0)%. The SP_3 model clearly outperforms the SP_1 model and yields 87.9 (78.1, 100.0)% sensitivity and 92.9 (84.6, 100.0)% specificity. Images computed with the ERT-based algorithm yield 90.9 (83.1, 100.0)% sensitivity and 97.6 (85.1, 100.0)% specificity.

We note that the upper bound of the CI for all models is 100.0%. The lower bound varies between models and between the sensitivity and the specificity. As in the case of the sensitivity and the specificity values, the CI of the ERT are smaller than those obtained with the SP_N models. However, the lower bounds of the SP_3 model are significantly higher than the lower bounds of the SP_1 model. The lower bound of the sensitivity with the SP_1 model is only 46.6% which is significantly lower than the 78.1% lower bound that is computed with the SP_3 model. Furthermore, the sensitivity computed with SP_1 images (66.7%) is lower than the lower bound of the sensitivity computed with the SP_3 model (78.1%). Similar results can be observed when comparing the specificity values.

Overall, these results indicate that the SP_3 -based reconstruction algorithm provides computational advantages over the ERT-based algorithm without sacrificing significant classification accuracy. In contrast, the SP_1 model provides computational advantages compared to the ERT at the expense of classification accuracy.

The names of the optimal features selected during the training phase and used in the testing phase are presented in Tables 10.3. The mean and standard error of the optimal features corresponding to the ERT and SP_3 models are plotted in Fig. 10.7.

The three optimal features chosen using the SP_3 model are F01:SV:a, F02:ST:a, and F13:VT:a. The first two features are the maximum and minimum values of the three dimensional data and the

Table 10.3: Name of optimal features selected during the training phase of the k -fold method using SP_N and ERT based reconstructions. Refer to Section 6.3.4 for a detailed explanation of the naming convention adopted for referencing image features.

Model	Name of Optimal Features
SP1	F01:ST:a, F04:GT:a, F34:SV:a, F16:VS:a, F03:SV:s, F04:VS:s, F05:VS:s, F04:VT:s
SP3	F01:SV:a, F02:ST:a, F26:VT:a
ERT	F01:UV:a, F02:SV:a, F05:SV:a, F02:ST:a, F08:GS:s

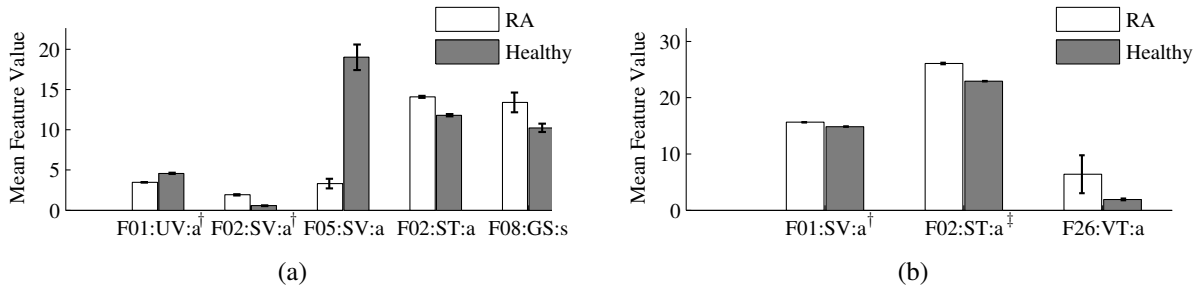


Figure 10.7: Mean value and standard error of the optimal features selected from analysis of the (a) ERT and (b) SP_3 based reconstruction images. A two-sample *student-t* test shows that differences between features from subjects with RA and without RA are statistically significant at the $\alpha = 0.01$ level. (a) Features denoted by \dagger are scaled by a factor of 10. (b) Features denoted by \dagger and \ddagger are scaled by a factor of 50 and 500, respectively.

sum of transverse slices, respectively (refer to Section 6.3.4 for details on the naming convention). The third feature corresponds to the largest coefficient of the Fourier transform of the variance between transverse slices.

The five optimal features chosen using ERT-based reconstruction images are F01:UV:a, F02:SV:a, F05:SV:a, F02:ST:a, and F08:GS:s. The first three features are the maximum, minimum, and ratio values of the three dimensional data, respectively. The fourth feature is the minimum of the sum of transverse slices. The fifth feature is a Gaussian mixture model (GMM) parameter.⁵

Together, the optimal features fit the general pattern first discussed in Chapter 7. While the optimal features are not the same features identified in Chapter 7, the general conclusion is still the same, primarily that the optimal features are associated with global absolute values of the

⁵More specifically, it is the second eigenvalue of the dominant positive Gaussian

absorption and scattering data and their spatial variation within the imaged section of the finger. The difference in optimal features is attributed to the difference in the rule for selecting the top 30 features and in the use of the k -fold method as an additional cross-validation step.

10.5 Discussion

Modifications to the objective function presented in Section 10.1 and the subsequent changes to the reconstruction algorithm are necessary when considering clinical data as the input measurements. The primary change that results from modification of the objective function is to the right hand side of the adjoint equation. Reconstruction with the PDE-constrained SP_1 and SP_3 algorithms is performed on a set of 219 human PIP joints, with 99 joints belonging to subjects with RA and 120 joints belonging to healthy subjects.

Computationally, the PDE-constrained SP_3 algorithm performs as expected, reaching convergence in less than 15 minutes. We note that compared to phantom studies, the algorithm is unable to reduce the objective function by several orders of magnitude. Instead, the algorithm settles on a local minimum that yields an objective function typically between 1.0% and 10.0% of the original objective function value. This difference in algorithm performance between phantom studies and the application to clinical data occurs because of the changes to the objective function, primarily due to the normalization of the input measurements by the average overall measured intensity. The normalization process is necessary as we are often not able to accurately specify the input source power and has the unfortunate effect of causing some loss of information provided by the input measurements.

Regardless of the challenges associated with using clinical data, the reconstruction algorithm performs as expected. Qualitatively, the reconstruction images of absorption and scattering meet our expectations. There appears to be a clear difference between joints of subjects with RA and subjects without RA. Primarily, joints from subjects with RA appear to have regions of elevated absorption and scattering. In contrast, these same regions in the joints of healthy individuals appear to have lower absorption and scattering values.

Corroborating one of the central theories of this thesis, we show that the sensitivity and specificity values computed with ERT-based images are higher than the values computed with the SP_3 -based images and even higher when compared to values obtained with SP_1 -based images. Additionally, the sensitivity and specificity values computed with SP_1 -based images are extremely poor

and may be of little clinical value. However, the sensitivity and specificity values computed with SP_3 -based images remain relatively high and may be sufficient for use in a clinical setting. Thus, the SP_3 model is clearly preferred over the SP_1 based on classification performance. From a computational efficiency standpoint, the differences between the SP_1 and SP_3 models is not significant.

The SP_3 models also compares favorably to the ERT model. While, both models achieve clinically relevant sensitivity and specificity values that compare favorably to values that can be expected from analysis of MRI and US images, it must be noted that the ERT-based images allow for very high sensitivities and specificities as compared to the SP_3 -based images. However, the SP_3 model requires significantly less computational resources compared to the ERT model. For example, in a typical high-performance desktop computer (such as a 6-core Mac Pro with 12 GB of RAM), one can easily perform the reconstruction of six finger joints (i.e. PIP joints II to IV of both hands of one subject) in less than 15 minutes, as all six reconstructions can be performed simultaneously (i.e. one reconstruction per core). In contrast, only one ERT-based reconstruction can be performed on the same computer, suggesting that a total of over 18 hours would be needed to complete the reconstruction of joints associated with a single subject.

Thus, a compelling argument can be made for favoring the SP_3 model over the ERT model as they both achieve clinically significant sensitivity and specificities, however, using the SP_3 model requires significantly less computational resources.

We note that the reconstruction maps of absorption and scattering obtained with the SP_N model do not, in general, match the images obtain with the ERT model. However, both sets of images fit into our understanding of what the true distribution of optical properties inside the imaged joint should be. This is generally expected as small differences in the algorithms can often lead to significant differences in the resulting reconstruction. This phenomena is due to the severe ill-posedness of the DOT inverse problem, where small differences in the optimization algorithm can results in the solution converging to any of the various local minima of the objective function.

Conclusion

The focus of this dissertation was to help further establish diffuse optical tomography (DOT) imaging as a valuable and clinically relevant tool. The work presented here reports on the development and implementation of computational algorithms that enhance the efficiency of the image reconstruction process and allow for the diagnosis of rheumatoid arthritis (RA) from analysis of DOT images.

11.1 Diagnosis of RA with CAD techniques

This dissertation first presented detailed analysis of frequency-domain DOT data from a clinical study where 36 subjects with RA were recruited, along with 20 healthy control subjects. The peripheral interphalangeal (PIP) joints II-IV of the dominant hand of each subject with RA was imaged, while PIP joints II-IV on both hands of the control subjects were also imaged. A total of 228 PIP joints were scanned during the study. Each joint was scanned with modulation frequencies of 600, 300, 0 MHz. The absorption and scattering coefficient maps of each joint were computed using a DOT reconstruction algorithm that uses the equation of radiative transfer (ERT) to model light propagation in biological tissue.¹

¹While a total of 36 subjects with RA were scanned, it must be noted that data from three subjects was discarded due to errors that occurred during the imaging procedure that rendered the DOT data unusable. Thus only 99 joints of subjects with RA were used in the analysis, resulting in a total of 219 joints (120 of the joints were from healthy

This set of PIP joint reconstruction-images formed the basis of the subsequent analysis, with a focus on understanding our ability to diagnose RA from these DOT images. The ability to diagnose RA was quantified through classification of each image as affected or not affected by RA using linear discriminate analysis (LDA) and leave- n -out cross validation.² Sensitivity (Se) and specificity (Sp) values were computed to quantify the ability to accurately diagnose RA.

The first set of results established that significantly higher sensitivities and specificities can be obtained from analysis of images obtained from 600 MHz (Se = 91.0%, Sp = 86.0%) and 300 MHz (Se = 81.0%, Sp = 74.0%) data than from 0 MHz (Se = 64.0.0%, Sp = 55.0.0%) data. The second set of results showed that joint classification using images computed with 0 MHz data were poor and, therefore, 0 MHz data was unlikely to play a significant role in the diagnostic process.

The third set of results showed that joints of subjects with RA that do not have detectable symptoms of RA on MRI and US scans,³ and therefore can falsely appear as healthy joints, are clearly different from healthy joints when comparing DOT images.⁴ Furthermore, these joints may also differ from joints with detectable symptoms on MRI and US images.

The fourth set of results showed for the first time that scattering image biomarkers were of significant importance. This is important as it had been previously reported that only absorption image features led to clinically relevant sensitivities and specificities. Features derived from the absorption coefficient image yield better one-dimensional classification results (Se = 78.0%, Sp = 78.0%) compared to features from the scattering coefficient images (Se = 75.0%, Sp = 75.0%). However, features from the scattering coefficient images yield significantly better performance

subjects).

²To be precise, the images themselves were not used. Instead, basic statical parameters were computed from each image and these features were subsequently used in the analysis. The image features included the maximum, minimum, variance, and ratio of maximum over the minimum of the spatial distribution of each optical parameter.

³The symptoms under consideration were erosion, effusion, and synovitis. Joints of subjects with RA that do not exhibit any of these symptoms are called asymptomatic PIP joints and they would be classified as not exhibiting symptoms of RA and the subjects may therefore be misdiagnosed if the diagnosis is made based on analysis of MRI or US images.

⁴ANOVA analysis was used to verify that joints of subjects with RA were statistically different from joints of subjects without RA at the 95.0% level. The same analysis showed that asymptomatic joints were not always statistically different from joints with RA, but were generally statistically different from healthy joints.

(Se = 83.0%, Sp = 83.0%) compared to the absorption coefficient images (Se = 91.0%, Sp = 86.0%) when multidimensional classification is considered.⁵

These findings are important as the study was the first to demonstrate that images of joints with RA can be successfully discriminated from image of joints without RA with significant accuracy. Furthermore, this study was the first to present conclusive evidence of the superiority of frequency domain imaging over steady-state domain imaging in the diagnosis of RA. Additionally, the study showed the importance of scattering coefficient images that had not been previously demonstrated. Those results inspired studies that evaluated the ability to diagnose RA using a combination of absorption and scattering coefficient features.

Together, these results showed that accurate classification of joints as affected or not affected with RA from analysis of DOT images can only be done when imaging is performed with a instrument that operates in the frequency domain, preferably at 600 MHz. Furthermore, it was shown that both the absorption and scattering coefficient images provide important pathological information that can aid in accurately classifying each joint.

Indeed, these results inspired the development of a computer-aided diagnosis (CAD) algorithm for automatically mining DOT data for image biomarkers that correlate with the presence of RA. Going beyond the initial analysis, the CAD algorithm was designed to take as input a set of DOT reconstruction data, extract a large set of image features that went well beyond the initial set of basic features, and mine the resulting set of image features for optimal classifiers that yield optimal or near optimal classification results.⁶

The CAD algorithm works as follows. Upon completion of the image reconstruction on unstructured tetrahedral meshes, the data is first converted into a Cartesian mesh for better analysis. Then, pre-processing of the reconstruction data is performed. This includes the computation of 9 two-dimensional data sets that are formed by projecting the three-dimensional structured data set

⁵All results presented here were obtained from analysis of images computed from data obtained with 600 MHz modulation frequency.

⁶The feature selection algorithm takes as input the 594 features extracted from each joint image and determines a subset of those features that yield the best diagnostic results.

along physiological planes.⁷ Features are then extracted from each of the 11 data sets.⁸

Three types of features are extracted from each of the 11 data sets. First, basic statistical parameters are computed, including the maximum, minimum, mean, and variance. The second type of features are Gaussian mixture models (GMM) coefficients.⁹ The third type of image features are the coefficient of the Fourier transform of each of the structured data sets.¹⁰ Overall, 297 features are extracted from each of the absorption and scattering coefficient images, resulting in 597 features for each joint.

The CAD algorithm then performs one-dimensional ROC analysis of each of the 597 features and selects the 30 features with the highest Youden index.¹¹ These 30 features are then used by the optimal feature selection algorithm to select a smaller subset of features that yield near optimal classification results. The algorithm was designed to use linear discriminate analysis (LDA), quadratic discriminate analysis (QDA), *k*-nearest neighbors (KNN), self-organizing maps (SOM), or support vector machines (SVM) to generate a decision rule for classifying an image as “affected” or “not affected” by RA.

The output of the algorithm is the optimal classifier and its performance metrics. The optimal classifier includes the optimal features and the accompanying classification algorithm. Performance is measured by the sensitivity, the specificity, and their corresponding 95.0% confidence intervals. The confidence intervals are particularly important because they account for any bias that may be introduced into the analysis by correlation between joints of the same subject.¹² Robust leave-one-subject-out cross-validation was used throughout the algorithm to ensure the integrity of the results.

⁷Projection planes are defined to coincide with each of the three physiological planes. Three projection operators were used. The operators sum all slices, compute the variance between all slices, or compute the average between all slices along each direction.

⁸The 11 data sets include the unstructured reconstruction data, the structured reconstruction data, and the 9 two-dimensional images computed by projecting the structured three-dimensional data.

⁹GMMs are a collection of Gaussian functions and are used to fit the spatial distribution of each optical parameter. The parameters that define the GMM accurately describe the distribution and location of volumes of high or low absorption (or scattering) coefficients in the image.

¹⁰Note that the Fourier transform was not applied to the unstructured data set.

¹¹The Youden index is defined as Sensitivity + Specificity - 1.

¹²The confidence interval is computed using the so-called effective sample size (ESS) and intra-class correlation coefficient (ICC).

The algorithm was tested using the same clinical data mentioned above, using only the 600 MHz data. We discovered that classification with SVM always performs better than classification with LDA, QDA, KNN, and SOM. Furthermore, for the SVM algorithm, a polynomial of order 3 was typically the optimal kernel.¹³ Finally, 100% (96.4%, 100.0%) sensitivity and 97.8% (93.8%, 100.0%) specificity values were achieved with an SVM classifier using a third-order polynomial kernel. The optimal image features were generally seen to capture the variation in optical parameters within a PIP joint.

11.2 Image reconstruction algorithm

The issue of inefficient image reconstruction was addressed by showing that an algorithm based on the SP_N light propagation model offers sufficient accuracy and is significantly more efficient than an algorithm based on the ERT model. To achieve the necessary improvements in reconstruction speed, a reduced-Hessian sequential quadratic programming (rSQP) PDE-constrained optimization approach was explored.

The derivation of the finite-volume discretization of the frequency-domain SP_N model, for $N = 1$ or 3 , was presented in detail. For the first time, we identified two competing versions of the frequency-domain SP_N model, a subject that has not been reported on previously. Algorithms for solving the two competing models were implemented, using the GMRES algorithm to solve the large-scale linear problem. Simulation results from each model were used to compare the accuracy and computational performance of each model. Each algorithm was validated through extensive numerical simulations. We discovered that the frequency-domain SP_N model which has only real-value diffusion-like coefficients is most accurate in situations where the optical properties more closely resemble tissue-like properties. As expected, the computational efficiency of the SP_N model compared to the ERT model was significant, performing up to 76 times faster than the ERT model. Additionally, the RAM required to solve the forward problem on a dense computational

¹³A low order polynomial kernel is indeed ideal as it helps to ensure the data is not over-fit by the model.

mesh with the SP_N model was approximately 315.1 MB, while solving the forward problem with the ERT model on the same phantom required over 6.15 GB of RAM.

Then, the necessary mathematical formulations for performing DOT reconstructions using the finite-volume frequency-domain SP_N model with a PDE-constrained optimization approach was presented. For comparison, the formulation of the traditional (“unconstrained”) reconstruction algorithm for the SP_N model was also developed and presented. A subroutine based on the limited-memory BFGS algorithm was developed to approximate and update the Hessian matrix of the quadratic problem associated with performing the image reconstruction. The algorithms were validated and their performance evaluated through extensive numerical simulations. The PDE-constrained approach was significantly faster than the unconstrained approach, often performing over one order of magnitude faster.

This was the first time that the finite-volume SP_N model was used to perform image reconstructions in DOT. This was also the first time that the SP_N model was used to perform reconstructions with a PDE-constrained optimization approach.

Finally, the rSQP PDE-constrained algorithm was used to reconstruct the absorption and scattering coefficient values of the finger joint clinical data. Reconstructions were performed with the $N = 1$ (i.e the diffusion equation) and the $N = 3$ models using the 600 MHz data. This was the first time reconstructions of frequency domain DOT data of finger joints were performed with light propagation models other than the ERT.

Analysis of the reconstruction data showed that subjects with RA had elevated absorption and scattering values in the location of the PIP joint compared to the images of healthy subjects. However, the images obtained with the SP_N -based algorithm differed from those obtained with the ERT-based algorithm. As both algorithms have been validated, it is difficult to definitively say which of the images are more correct as the true distribution of optical parameters in each joint is not known. It is encouraging, nevertheless, that the absorption and scattering coefficient distributions match our expectations (based on our understanding of the physiological processes and symptoms associated with the onset of RA).

To quantitatively compare the utility of the SP_N and ERT-based images, the CAD algorithm was used to process the resulting reconstruction data and to quantify the ability to diagnose RA using these SP_N -based images. Clinically significant sensitivity and specificity values were achieved when using images obtained using the SP_3 model, while images obtained with the SP_1 model resulted in significantly lower sensitivity and specificity values. While images obtained with the SP_3 allowed for clinically relevant sensitivity and specificity values, it was clear that reconstruction images obtained using the ERT model allowed for even higher sensitivity and specificity values. An additional step was taken to ensure the integrity of the classification results. In a blinded process, the data was divided into training and testing sets. The training set was used as input to the CAD algorithm, from which optimal classifiers were computed. Then, the testing set was used to compute the sensitivity and specificity with the selected classifier.¹⁴

The classification performance achieved with ERT-based images was used as a benchmark, as analysis of these images resulted in 90.9% (77.1%, 100.0%) sensitivity and 97.6% (81.1%, 100.0%) specificity.¹⁵ The SP_3 model performed relatively well, with 87.9% (69.4%, 100.0%) sensitivity and 92.9% (71.7%, 100.0%) specificity. The performance of the diffusion model (SP_1) was the worst, with 66.7% (46.6%, 100.0%) sensitivity and 81.1% (64.8%, 100.0%) specificity.

From a computational standpoint, however, the difference in resources required to solve a typical inverse problem with the ERT model (≥ 6.0 GB) was significantly more than the resources required to solve the same inverse problem with the SP_3 model (≤ 200.0 MB). The low computational resources required by the SP_3 model imply that multiple reconstructions can be performed simultaneously on standard desktop computers. This is advantageous when the goal is to produce a clinically useful tool. For example, the rSQP PDE-constrained SP_N algorithm can reconstruct the absorption and scattering coefficient images of a single subject in less than 15.0 minutes on a standard desktop computer. In contrast, using a highly efficient ERT algorithm would require at

¹⁴This is the so-called k -fold cross-validation procedure. The training set consisted of 2/3 of the data and the testing set contained the remaining 1/3 of data.

¹⁵The numbers in parentheses represent the 95.0% confidence interval for the reported diagnostic value. The confidence interval incorporates any potential correlation that may exist between finger joints of the same subject and corrects for that bias. Correlation between joints results in decreasing the effective sample size and therefore widens the confidence interval.

least 18.0 hours on the same computer.¹⁶ The advantages are clear, as the SP_N based algorithm would allow the generation of finger joint images that could be analyzed almost immediately after scanning the subject.

Overall, we presented sufficient evidence to establish that the SP_3 light propagation model provides sufficiently accurate DOT reconstruction images to achieve clinically significant diagnostic results with a significant reduction in computation time and system resources.

11.3 Future work

A prospective study involving a large number of subjects is necessary to further validate the diagnostic results presented in this dissertation. For example, the top features discovered in Chapter 7 and in Chapter 10 can be used along with the SVM classifier to diagnose future subjects. Such a prospective study will offer sufficient insight into whether RA can or cannot be diagnosed from DOT images. Furthermore, the identified classifiers can be used to identify subjects with RA that respond to therapy and those that do not respond. Indeed, such a study has already been designed and is under consideration by our team of collaborators. The study is designed to be a longitudinal study where response to therapy will also be monitored. Based on power calculations, it was established that 110 subjects would provide sufficient statistical power to accurately identify responders from non-responders to therapy.¹⁷

There are three additional areas that I believe should be the focus of future work that builds on the results reported in this dissertation. These include the application of the algorithms presented in this dissertation to additional therapeutic areas, refinement of these fragmented algorithms into a clinically useful all-in-one package, and further enhancement of the SP_N reconstruction algorithm through the use of parallel computing techniques.

¹⁶Here we assume that a typical desktop computer is a machine with at least a six cores and 6 GB of RAM. We also assume that a given subject has three joints imaged per hand.

¹⁷The effective sample size is actually much larger than 110 as all 10 fingers of each subjects will be imaged, resulting in 1,100 distinct joints. The effective sample size, taking into consideration the correlation between joints of the same subjects, is expected to be approximately between 400 and 700.

First, a logical extension of the computer-aided diagnosis work is to focus on the application of the feature selection techniques used throughout this dissertation to clinical data associated with the diagnosis and monitoring of peripheral arterial disease (PAD) and breast cancer. Indeed, preliminary results from the use of the CAD algorithm to time-domain DOT data from a small clinical study involving subjects with PAD shows that sensitivity and specificity values in the range of 80% to 90% can be achieved. Modification to the algorithm as presented in this dissertation include the extraction of heuristic temporal features to augment the spatial features already discussed. In breast imaging, the CAD algorithm can be used to discover spatial-temporal features that correlate with the presence of breast cancer. These DOT image features can also be used to predict and monitor response to therapy in the neoadjuvant chemotherapy setting. As in the case of RA, retrospective studies can be used to determine optimal image biomarkers and prospective studies can be used to validate their utility.

Similarly, the PDE-constrained SP_N reconstruction algorithm can be a valuable tool for use in clinical studies that involve large data sets, such as PAD and breast imaging, where often over 1,000 three-dimensional “frames” are captured and must be reconstructed. Using ERT-based reconstruction schemes is impractical because of the associated computational burden and using the diffusion model introduces severe modeling errors. However, the SP_3 model is not more expensive to use with clinical data than the diffusion model and yields solutions that can more accurately approximate ERT-based reconstructions.

Second, a holistic approach to the integration of DOT into the clinical setting should include an algorithm that directly integrates into the imaging hardware and is capable of taking as input the “raw” imaging data and automatically output DOT images and a diagnosis recommendation. This would require integrating the reconstruction algorithm and the CAD algorithm, as they individually are already capable of providing the necessary results. This would help further establish DOT as a valuable tool in the clinical setting as it would be user-friendly.

Third, the reconstruction process can be further enhanced so that images can be obtained within one to two minutes, providing almost immediate results that can be easily analyzed at the point of

imaging, without sacrificing significant accuracy. This can be achieved by reducing the reconstruction time associated with executing the PDE-constrained SP_N algorithm through the use of parallel computing techniques. For example, a typical DOT imaging scan uses multiple sources (11 sources in the case of PIP joints) and records each source with all detectors. Recovering the optical parameter distribution requires solving the forward problem at least twice during the reconstruction process and the adjoint problem at least once per inverse iteration. However, each of these problems can be decomposed into smaller linear problems that are associated with only one source at a time. As such, using at least 11 processors allows one to distribute the work associated with a single source (i.e. the forward or adjoint problem) to a specific processor. Then, the time necessary to solve these smaller linear problems for each of the 11 sources can be reduced to the time needed to solve for one source.¹⁸ This strategy is becoming increasingly feasible as desktop computers with 12 and 16 CPU cores are becoming increasingly common and inexpensive.

11.4 Discussion

The field of diffuse optical tomography has progressed significantly over the past 20 years and is now close to being established as a valuable tool at the clinical level, having already shown great value as a pre-clinical tool. To ensure the continual progress of this novel imaging modality it is imperative that researchers continue to formulate and execute clinical studies and apply computational and statistical techniques to this data to further demonstrate the diagnostic utility of DOT. Large prospective clinical studies are needed to effectively and irrefutably show what there are disease specific image biomarkers that can be used to accurately diagnose and monitor a multitude of diseases, including RA, breast cancer, peripheral arterial disease, and infantile hemangiomas, among others. We have taken but a first step towards that goal with the work presented in this dissertation, and undoubtedly, the case for DOT as a vital clinical tool will be further strengthened over the next five to ten years as results from additional studies are published.

¹⁸A near-linear speedup is expected with this type of domain decomposition as the associated communication between processors is minimal.

Chapter 12

Bibliography

- [1] A. D. Klose and E. W. Larsen, “Light transport in biological tissue based on the simplified spherical harmonics equations,” *Journal of Computational Physics*, vol. 220, no. 1, pp. 441 – 470, 2006.
- [2] F. C. Arnett, S. M. Edworthy, D. A. Bloch, D. J. McShane, J. F. Fries, N. S. Cooper, L. A. Healey, S. R. Kaplan, M. H. Liang, and H. S. Luthra, “The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis,” *Arthritis Rheum.*, vol. 31, pp. 315–324, Mar 1988.
- [3] S. A. Prahl, “Optical properties spectra.” As compiled and posted on <http://omlc.ogi.edu/spectra/>, last accessed on June 1, 2013.
- [4] W. B. Gratzer, “Molar extinction coefficient for hemoglobin in water.” As compiled and posted on <http://omlc.ogi.edu/spectra/hemoglobin/summary.html> by S. A. Prahl, last accessed on June 1, 2013.
- [5] N. Kollias, “Molar extinction coefficient for hemoglobin in water.” As compiled and posted on <http://omlc.ogi.edu/spectra/hemoglobin/summary.html> by S. A. Prahl, last accessed on June 1, 2013.
- [6] R. L. van Veen, H. Sterenborg, A. Pifferi, A. Torricelli, and R. Cubeddu, “Determination of vis-nir absorption coefficients of mammalian fat, with time- and spatially resolved diffuse reflectance and transmission spectroscopy,” in *Biomedical Topical Meeting*, p. SF4, Optical Society of America, 2004.

- [7] D. Yim, "The specific absorption coefficient of water." As compiled and posted on <http://www.npsg.uwaterloo.ca/data/blood.php> by Natural Phenomena Simulation Group, last accessed on June 1, 2013.
- [8] K. F. Palmer and D. Williams, "Optical properties of water in the near infrared," *J. Opt. Soc. Am.*, vol. 64, pp. 1107–1110, Aug 1974.
- [9] A. Hielscher, H. K. Kim, L. Montejo, S. Blaschke, U. Netz, P. Zwaka, G. Illing, G. Muller, and J. Beuthan, "Frequency-domain optical tomographic imaging of arthritic finger joints," *Medical Imaging, IEEE Transactions on*, vol. 30, pp. 1725 –1736, oct. 2011.
- [10] A. H. Hielscher, A. D. Klose, A. K. Scheel, B. Moa-Anderson, M. Backhaus, U. Netz, and J. Beuthan, "Sagittal laser optical tomography for imaging of rheumatoid finger joints," *Physics in Medicine and Biology*, vol. 49, no. 7, p. 1147, 2004.
- [11] B. J. Tromberg, B. W. Pogue, K. D. Paulsen, A. G. Yodh, D. A. Boas, and A. E. Cerussi, "Assessing the future of diffuse optical imaging technologies for breast cancer management," *Medical Physics*, vol. 35, no. 6, pp. 2443–2451, 2008.
- [12] Q. Zhang and H. Jiang, "Three-dimensional diffuse optical imaging of hand joints: System description and phantom studies," *Optics and Lasers in Engineering*, vol. 43, no. 11, pp. 1237 – 1251, 2005.
- [13] A. K. Scheel, M. Backhaus, A. D. Klose, B. Moa-Anderson, U. J. Netz, K.-G. A. Hermann, J. Beuthan, G. A. Müller, G. R. Burmester, and A. H. Hielscher, "First clinical evaluation of sagittal laser optical tomography for detection of synovitis in arthritic finger joints," *Annals of the Rheumatic Diseases*, vol. 64, pp. 239–245, August 2005.
- [14] J. M. Lasker, C. J. Fong, D. T. Ginat, E. Dwyer, and A. H. Hielscher, "Dynamic optical imaging of vascular and metabolic reactivity in rheumatoid joints," *Journal of Biomedical Optics*, vol. 12, no. 5, p. 052001, 2007.
- [15] Q. Fang, S. Carp, J. Selb, G. Boverman, Q. Zhang, D. Kopans, R. Moore, E. Miller, D. Brooks, and D. Boas, "Combined optical imaging and mammography of the healthy breast: Optical contrast

- derived from breast structure and compression,” *Medical Imaging, IEEE Transactions on*, vol. 28, pp. 30–42, Jan. 2009.
- [16] S. R. Arridge, “Optical tomography in medical imaging,” *Inverse Problems*, vol. 15, no. 2, p. R41, 1999.
- [17] S. R. Arridge and J. C. Schotland, “Optical tomography: forward and inverse problems,” *Inverse Problems*, vol. 25, no. 12, p. 123010, 2009.
- [18] T. Durduran, R. Choe, W. B. Baker, and A. G. Yodh, “Diffuse optics for tissue monitoring and tomography,” *Reports on Progress in Physics*, vol. 73, no. 7, p. 076701, 2010.
- [19] A. D. Klose, “Radiative transfer of luminescence light in biological tissue,” in *Light Scattering Reviews 4* (A. A. Kokhanovsky, ed.), Springer Praxis Books, pp. 293–345, Springer Berlin Heidelberg, 2009.
- [20] H. K. Kim and A. H. Hielscher, “A pde-constrained sqp algorithm for optical tomography based on the frequency-domain equation of radiative transfer,” *Inverse Problems*, vol. 25, no. 1, p. 015010, 2009.
- [21] H. K. Kim, J. H. Lee, and A. Hielscher, “Pde-constrained fluorescence tomography with the frequency-domain equation of radiative transfer,” *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, no. 4, pp. 793–803, 2010.
- [22] A. Gibson and H. Dehghani, “Diffuse optical imaging,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1900, pp. 3055–3072, 2009.
- [23] W. Bangerth and A. Joshi, “Adaptive finite element methods for the solution of inverse problems in optical tomography,” *Inverse Problems*, vol. 24, no. 3, p. 034011, 2008.
- [24] M. Bartels, W. Chen, R. Bardhan, S. Ke, N. Halas, T. Wareing, J. McGhee, and A. Joshi, “Multimodal optical molecular image reconstruction with frequency domain measurements,” in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 6655–6658, Sept. 2009.

- [25] A. Joshi, J. C. Rasmussen, E. M. Sevick-Muraca, T. A. Wareing, and J. McGhee, “Radiative transport-based frequency-domain fluorescence tomography,” *Physics in Medicine and Biology*, vol. 53, no. 8, p. 2069, 2008.
- [26] S. Wright, M. Schweiger, and S. R. Arridge, “Reconstruction in optical tomography using the p n approximations,” *Measurement Science and Technology*, vol. 18, no. 1, p. 79, 2007.
- [27] G. Bal, “Reconstructions in impedance and optical tomography with singular interfaces,” *Inverse Problems*, vol. 21, no. 1, p. 113, 2005.
- [28] A. Karellas and S. Vedantham, “Breast cancer imaging: A perspective for the next decade,” *Medical Physics*, vol. 35, no. 11, pp. 4878–4897, 2008.
- [29] J. Z. Wang, X. Liang, Q. Zhang, L. L. Fajardo, and H. Jiang, “Automated breast cancer classification using near-infrared optical tomographic images,” *Journal of Biomedical Optics*, vol. 13, no. 4, pp. 044001–044001–10, 2008.
- [30] A. T. Eggebrecht, B. R. White, S. L. Ferradal, C. Chen, Y. Zhan, A. Z. Snyder, H. Dehghani, and J. P. Culver, “A quantitative spatial comparison of high-density diffuse optical tomography and fmri cortical mapping,” *NeuroImage*, vol. 61, no. 4, pp. 1120 – 1128, 2012.
- [31] S. Madsen, *Optical Methods and Instrumentation in Brain Imaging and Therapy*. Bioanalysis (En ligne), Springer New York, 2013.
- [32] A. V. Medvedev, J. Kainerstorfer, S. V. Borisov, R. L. Barbour, and J. VanMeter, “Event-related fast optical signal in a rapid object recognition task: Improving detection by the independent component analysis,” *Brain Research*, vol. 1236, pp. 145 – 158, 2008.
- [33] T. J. Huppert, S. G. Diamond, and D. A. Boas, “Direct estimation of evoked hemoglobin changes by multimodality fusion imaging,” *Journal of Biomedical Optics*, vol. 13, no. 5, p. 054031, 2008.
- [34] A. K. Dunn, H. Bolay, M. A. Moskowitz, and D. A. Boas, “Dynamic imaging of cerebral blood flow using laser speckle,” *Cerebral Blood Flow and Metabolism*, vol. 21, pp. 195–201, 2001.

- [35] Z. Yuan, Q. Zhang, E. Sobel, and H. Jiang, "Three-dimensional diffuse optical tomography of osteoarthritis: initial results in the finger joints," *Journal of Biomedical Optics*, vol. 12, no. 3, p. 034001, 2007.
- [36] A. D. Klose, V. Prapavat, O. Minet, J. Beuthan, and G. J. Mueller, "Ra diagnostics applying optical tomography in frequency domain," *SPIE Proc., Optical and Imaging Techniques for Biomonitoring III*, vol. 3196, no. 1, pp. 194–204, 1998.
- [37] A. D. Klose, A. H. Hielscher, K. M. Hanson, and J. Beuthan, "Two- and three-dimensional optical tomography of finger joints for diagnostics of rheumatoid arthritis," *Photon Propagation in Tissues IV*, vol. 3566, no. 1, pp. 151–160, 1998.
- [38] D. Aletaha, T. Neogi, A. J. Silman, J. Funovits, D. T. Felson, C. O. Bingham, N. S. Birnbaum, G. R. Burmester, V. P. Bykerk, M. D. Cohen, B. Combe, K. H. Costenbader, M. Dougados, P. Emery, G. Ferraccioli, J. M. Hazes, K. Hobbs, T. W. Huizinga, A. Kavanaugh, J. Kay, T. K. Kvien, T. Laing, P. Mease, H. A. Menard, L. W. Moreland, R. L. Naden, T. Pincus, J. S. Smolen, E. Stanislawska-Biernat, D. Symmons, P. P. Tak, K. S. Upchurch, J. Vencovsky, F. Wolfe, and G. Hawker, "2010 rheumatoid arthritis classification criteria: an american college of rheumatology/european league against rheumatism collaborative initiative," *Annals of the Rheumatic Diseases*, vol. 69, no. 9, pp. 1580–1588, 2010.
- [39] F. Banal, M. Dougados, C. Combescure, and L. Gossec, "Sensitivity and specificity of the american college of rheumatology 1987 criteria for the diagnosis of rheumatoid arthritis according to disease duration: a systematic literature review and meta-analysis," *Annals of the Rheumatic Diseases*, vol. 68, no. 7, pp. 1184–1191, 2009.
- [40] P. Libby, P. M. Ridker, and A. Maseri, "Inflammation and atherosclerosis," *Circulation*, vol. 105, no. 9, pp. 1135–1143, 2002.
- [41] J. Beutel and H. Kundel, *Handbook of Medical Imaging: Physics and psychophysics*. No. v. 2 in Handbook of Medical Imaging, SPIE Press, 2000.
- [42] S. L. Jacques, "Optical properties of biological tissues: a review," *Physics in Medicine and Biology*, vol. 58, no. 11, p. R37, 2013.

- [43] R. Weissleder and V. Ntziachristos, "Shedding light onto live molecular targets," *Nat. Med.*, vol. 9, pp. 123–128, Jan 2003.
- [44] G. C. Salzman, J. M. Crowell, J. C. Martin, T. T. Trujillo, A. Romero, P. F. Mullaney, and P. M. LaBauve, "Cell classification by laser light scattering: identification and separation of unstained leukocytes," *Acta Cytol.*, vol. 19, no. 4, pp. 374–377, 1975.
- [45] L. S. Cram and A. Brunsting, "Fluorescence and light-scattering measurements on hog cholera-infected PK-15 cells," *Exp. Cell Res.*, vol. 78, pp. 209–213, Mar 1973.
- [46] J. R. Mourant, M. Canpolat, C. Brocker, O. Esponda-Ramos, T. M. Johnson, A. Matanock, K. Stetter, and J. P. Freyer, "Light scattering from cells: the contribution of the nucleus and the effects of proliferative status," *J Biomed Opt*, vol. 5, pp. 131–137, Apr 2000.
- [47] B. Beauvoit, T. Kitai, and B. Chance, "Contribution of the mitochondrial compartment to the optical properties of the rat liver: a theoretical and practical approach," *Biophys. J.*, vol. 67, pp. 2501–2510, Dec 1994.
- [48] O. C. Marina, C. K. Sanders, and J. R. Mourant, "Correlating light scattering with internal cellular structures," *Biomed Opt Express*, vol. 3, pp. 296–312, Feb 2012.
- [49] J. R. Mourant, J. P. Freyer, A. H. Hielscher, A. A. Eick, D. Shen, and T. M. Johnson, "Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics," *Appl. Opt.*, vol. 37, pp. 3586–3593, Jun 1998.
- [50] C. Bohren and D. Huffman, *Absorption and Scattering of Light by Small Particles*. Wiley Science Series, Wiley, 2008.
- [51] R. Cotran, V. Kumar, and S. Robbins, *Robbins pathologic basis of disease*. Robbins Pathology Series, Saunders, 1994.
- [52] S. A. Prahl, M. Keijzer, S. L. Jacques, and A. J. Welch, "A monte carlo model of light propagation in tissue," *Proc. SPIE IS*, vol. 5, pp. 102–111, 1989.

- [53] L. Wang, S. L. Jacques, and L. Zheng, "Mcm1–monte carlo modeling of light transport in multi-layered tissues," *Computer Methods and Programs in Biomedicine*, vol. 47, no. 2, pp. 131 – 146, 1995.
- [54] Q. Fang and D. A. Boas, "Monte carlo simulation of photon migration in 3d turbid media accelerated by graphics processing units," *Opt. Express*, vol. 17, pp. 20178–20190, Oct 2009.
- [55] M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties," *Appl. Opt.*, vol. 28, pp. 2331–2336, Jun 1989.
- [56] S. R. Arridge, M. Cope, and D. T. Delpy, "The theoretical basis for the determination of optical pathlengths in tissue: temporal and frequency analysis," *Physics in Medicine and Biology*, vol. 37, no. 7, p. 1531, 1992.
- [57] S. R. Arridge, M. Schweiger, M. Hiraoka, and D. T. Delpy, "A finite element approach for modeling photon transport in tissue," *Medical Physics*, vol. 20, no. 2, pp. 299–309, 1993.
- [58] A. H. Hielscher, R. E. Alcouffe, and R. L. Barbour, "Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues," *Physics in Medicine and Biology*, vol. 43, no. 5, p. 1285, 1998.
- [59] O. Dorn, "A transport-backtransport method for optical tomography," *Inverse Problems*, vol. 14, no. 5, p. 1107, 1998.
- [60] A. D. Klose and A. H. Hielscher, "Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer," *Medical Physics*, vol. 26, no. 8, pp. 1698–1707, 1999.
- [61] 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 and Radiative Transfer*, vol. 72, no. 5, pp. 691 – 713, 2002.
- [62] K. Ren, G. S. Abdoulaev, G. Bal, and A. H. Hielscher, "Algorithm for solving the equation of radiative transfer in the frequency domain," *Opt. Lett.*, vol. 29, pp. 578–580, Mar 2004.

- [63] M. B. Salah, F. Askri, and S. B. Nasrallah, "Unstructured control-volume finite-element method for radiative heat transfer in a complex 2-d geometry," *Numerical Heat Transfer, Part B: Fundamentals: An International Journal of Computation and Methodology*, vol. 48, no. 5, pp. 477–497, 2005.
- [64] E. Alerstam, T. Svensson, and S. Andersson-Engels, "Parallel computing with graphics processing units for high-speed monte carlo simulation of photon migration," *Journal of Biomedical Optics*, vol. 13, no. 6, p. 060504, 2008.
- [65] K. Ren, G. Bal, and A. H. Hielscher, "Transport- and diffusion-based optical tomography in small domains: a comparative study," *Appl. Opt.*, vol. 46, pp. 6669–6679, Sep 2007.
- [66] L. D. Montejo, A. D. Klose, and A. H. Hielscher, "Implementation of the equation of radiative transfer on block-structured grids for modeling light propagation in tissue," *Biomed. Opt. Express*, vol. 1, pp. 861–878, Oct 2010.
- [67] 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.
- [68] K. M. Case and P. F. Zweifel, *Linear transport theory*. Addison-Wesley, Reading, 1967.
- [69] J. J. Duderstadt and W. R. Martin, *Transport Theory*. John Wiley & Sons, 1979.
- [70] E. E. Lewis and W. F. Miller, *Computational Methods of Neutron Transport*. Wiley, New York, 1984.
- [71] B. G. Carlson and K. D. Lathrop, *Transport theory - the method of discrete ordinates*. Gordon and Breach, New York, 1968.
- [72] K. Morton and D. Mayers, *Numerical Solution of Partial Differential Equations: An Introduction*. Numerical Solution of Partial Differential Equations: An Introduction, Cambridge University Press, 2005.
- [73] M. Chu, K. Vishwanath, A. D. Klose, and H. Dehghani, "Light transport in biological tissue using three-dimensional frequency-domain simplified spherical harmonics equations," *Physics in Medicine and Biology*, vol. 54, no. 8, p. 2493, 2009.

- [74] J. B. Domínguez and Y. Bérubé-Lauzière, “Diffuse light propagation in biological media by a time-domain parabolic simplified spherical harmonics approximation with ray-divergence effects,” *Appl. Opt.*, vol. 49, pp. 1414–1429, Mar 2010.
- [75] V. D. Liseikin, *Grid Generation Methods*. Springer Netherlands, 2010.
- [76] J. C. Rasmussen, A. Joshi, T. Pan, T. Wareing, J. McGhee, and E. M. Sevick-Muraca, “Radiative transport in fluorescence-enhanced frequency domain photon migration,” *Medical Physics*, vol. 33, no. 12, pp. 4685–4700, 2006.
- [77] X. Gu, K. Ren, and A. H. Hielscher, “Frequency-domain sensitivity analysis for small imaging domains using the equation of radiative transfer,” *Appl. Opt.*, vol. 46, pp. 1624–1632, Apr 2007.
- [78] A. D. Klose and A. Hielscher, “Fluorescence tomography with simulated data based on the equation of radiative transfer,” *Opt. Lett.*, vol. 28, pp. 1019–1021, Jun 2003.
- [79] A. D. Klose, B. J. Beattie, H. Dehghani, L. Vider, C. Le, V. Ponomarev, and R. Blasberg, “In vivo bioluminescence tomography with a blocking-off finite-difference sp³ method and mri/ct coregistration,” *Medical Physics*, vol. 37, no. 1, pp. 329–338, 2010.
- [80] M. P. M. van der Linden, R. Knevel, T. W. J. Huizinga, and A. H. M. van der Helm-van Mil, “Classification of rheumatoid arthritis: Comparison of the 1987 american college of rheumatology criteria and the 2010 american college of rheumatology/european league against rheumatism criteria,” *Arthritis & Rheumatism*, vol. 63, no. 1, pp. 37–42, 2011.
- [81] E. Myasoedova, C. S. Crowson, H. M. Kremers, T. M. Therneau, and S. E. Gabriel, “Is the incidence of rheumatoid arthritis rising?: Results from olmsted county, minnesota, 1955–2007,” *Arthritis & Rheumatism*, vol. 62, no. 6, pp. 1576–1582, 2010.
- [82] V. Majithia and S. A. Geraci, “Rheumatoid arthritis: Diagnosis and management,” *The American Journal of Medicine*, vol. 120, no. 11, pp. 936 – 939, 2007.
- [83] J. R. O’Dell, J. S. Smolen, D. Aletaha, D. R. Robinson, and E. W. St. Clair, “Rheumatoid arthritis,” in *A Clinician’s Pearls and Myths in Rheumatology* (J. H. Stone, ed.), pp. 1–13, Springer London, 2010.

- [84] T. Pincus, "The underestimated long term medical and economic consequences of rheumatoid arthritis," *Drugs*, vol. 50 Suppl 1, pp. 1–14, 1995.
- [85] D. L. Scott, F. Wolfe, and T. W. Huizinga, "Rheumatoid arthritis," *The Lancet*, vol. 376, no. 9746, pp. 1094 – 1108, 2010.
- [86] J. Norris, *Professional guide to diseases*. Professional Guide to Diseases Series, Springhouse Corporation, 1998.
- [87] C. G. Helmick, D. T. Felson, R. C. Lawrence, S. Gabriel, R. Hirsch, C. K. Kwoh, M. H. Liang, H. M. Kremers, M. D. Mayes, P. A. Merkel, S. R. Pillemer, J. D. Reveille, J. H. Stone, and N. A. D. Workgroup, "Estimates of the prevalence of arthritis and other rheumatic conditions in the united states: Part i," *Arthritis & Rheumatism*, vol. 58, no. 1, pp. 15–25, 2008.
- [88] E. Harris, "Rheumatoid arthritis. pathophysiology and implications for therapy," *New Engl J Med.*, vol. 322, pp. 1277–1289, 1990.
- [89] D. P. Lubeck, "The economic impact of arthritis," *Arthritis Care Res*, vol. 8, pp. 304–310, Dec 1995.
- [90] C. Wilder, "Disability days: United states, 1983," *Vital and Health Statistics*, no. 158, 1986.
- [91] D. M. Lee and M. E. Weinblatt, "Rheumatoid arthritis," *The Lancet*, vol. 358, no. 9285, pp. 903 – 911, 2001.
- [92] G. Rodnan, H. Schumacher, and A. Foundation, *Primer on the rheumatic diseases*. Arthritis Foundation, 1983.
- [93] J. Edeiken, *Roentgen diagnosis of diseases of bone*. Golden's diagnostic radiology series, Williams & Wilkins, 1981.
- [94] D. E. Furst, "The risk of infections with biologic therapies for rheumatoid arthritis," *Seminars in Arthritis and Rheumatism*, vol. 39, no. 5, pp. 327 – 346, 2010.
- [95] P. Laudicina, *Applied pathology for radiographers*. Saunders, 1989.
- [96] O. STEINBROCKER, C. H. TRAEGER, and R. C. BATTERMAN, "Therapeutic criteria in rheumatoid arthritis," *J Am Med Assoc*, vol. 140, pp. 659–662, Jun 1949.

- [97] B. Fautrel, “Diagnosing Early or Rheumatoid Arthritis. Which Is Better: Expert Opinion or Evidence?,” *JOURNAL OF RHEUMATOLOGY*, vol. 36, pp. 2375–2377, NOV 2009.
- [98] K. W. Chan, D. T. Felson, R. A. Yood, and A. M. Walker, “The lag time between onset of symptoms and diagnosis of rheumatoid arthritis,” *Arthritis Rheum.*, vol. 37, pp. 814–820, Jun 1994.
- [99] E. Solau-Gervais, J.-L. Legrand, B. Cortet, B. Duquesnoy, and R.-M. Flipo, “Magnetic resonance imaging of the hand for the diagnosis of rheumatoid arthritis in the absence of anti-cyclic citrullinated peptide antibodies: a prospective study,” *The Journal of Rheumatology*, vol. 33, no. 9, pp. 1760–1765, 2006.
- [100] H. D. El-Gabalawy and P. E. Lipsky, “Why do we not have a cure for rheumatoid arthritis?,” *Arthritis Res.*, vol. 4 Suppl 3, pp. 297–301, 2002.
- [101] A. Wluka, R. Buchbinder, A. Mylvaganam, S. Hall, A. Harkness, D. Lewis, G. O. Littlejohn, M. H. Miller, and P. F. Ryan, “Longterm methotrexate use in rheumatoid arthritis: 12 year followup of 460 patients treated in community practice,” *J. Rheumatol.*, vol. 27, pp. 1864–1871, Aug 2000.
- [102] D. Krause, B. Schleusser, G. Herborn, and R. Rau, “Response to methotrexate treatment is associated with reduced mortality in patients with severe rheumatoid arthritis,” *Arthritis Rheum.*, vol. 43, pp. 14–21, January 2000.
- [103] J. M. Bathon, R. W. Martin, R. M. Fleischmann, J. R. Tesser, M. H. Schiff, E. C. Keystone, M. C. Genovese, M. C. Wasko, L. W. Moreland, A. L. Weaver, J. Markenson, and B. K. Finck, “A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis,” *New England Journal of Medicine*, vol. 343, no. 22, pp. 1586–1593, 2000.
- [104] J. T. Sharp, V. Strand, H. Leung, F. Hurley, and I. Loew-Friedrich, “Treatment with leflunomide slows radiographic progression of rheumatoid arthritis: results from three randomized controlled trials of leflunomide in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group,” *Arthritis Rheum.*, vol. 43, pp. 495–505, Mar 2000.

- [105] J. F. Fries, C. A. Williams, D. Morfeld, G. Singh, and J. Sibley, "Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies," *Arthritis Rheum.*, vol. 39, pp. 616–622, Apr 1996.
- [106] M. E. Weinblatt, "Rheumatoid arthritis: Treat now, not later!," *Annals of Internal Medicine*, vol. 124, no. 8, pp. 773–774, 1996.
- [107] P. Fishman and S. Bar-Yehuda, "Rheumatoid arthritis: History, molecular mechanisms and therapeutic applications," in *A3 Adenosine Receptors from Cell Biology to Pharmacology and Therapeutics* (P. A. Borea, ed.), pp. 291–298, Springer Netherlands, 2010.
- [108] P. Tugwell, T. Pincus, D. Yocum, M. Stein, O. Gluck, G. Kraag, R. McKendry, J. Tesser, P. Baker, and G. Wells, "Combination therapy with cyclosporine and methotrexate in severe rheumatoid arthritis," *New England Journal of Medicine*, vol. 333, no. 3, pp. 137–142, 1995. PMID: 7791814.
- [109] P. Tugwell, G. Wells, V. Strand, A. Maetzel, C. Bombardier, B. Crawford, C. Dorrier, and A. Thompson, "Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis - sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial," *Arthritis & Rheumatism*, vol. 43, no. 3, pp. 506–514, 2000.
- [110] I. A. Jaffe, "New approaches to the management of rheumatoid arthritis," *J Rheumatol Suppl*, vol. 36, pp. 2–8, Nov 1992.
- [111] L. B. A. van de Putte and P. L. C. M. van Riel, "Rheumatoid arthritis-changing theories and treatment modalities," *Acta Orthopaedica*, vol. 66, no. s266, pp. 10–13, 1995.
- [112] A. Tedesco, D. D'Agostino, I. Soriente, P. Amato, R. Piccoli, and P. Sabatini, "A new strategy for the early diagnosis of rheumatoid arthritis: A combined approach," *Autoimmunity Reviews*, vol. 8, no. 3, pp. 233 – 237, 2009. Special Issue on Signal Processing by Immune System in SLE.
- [113] J. H. Brown and S. A. Deluca, "The radiology of rheumatoid arthritis," *Am Fam Physician*, vol. 52, pp. 1372–1380, Oct 1995.

- [114] M. Ostergaard and M. Szkudlarek, "Imaging in rheumatoid arthritis – why mri and ultrasonography can no longer be ignored," *Scandinavian Journal of Rheumatology*, vol. 32, no. 2, pp. 63–73, 2003.
- [115] W. A. Schmidt, M. Backhaus, H. Sattler, and H. Kellner, "Bildgebende verfahren in der rheumatologie: Sonographie bei rheumatoider arthritis," *Zeitschrift für Rheumatologie*, vol. 62, pp. 23–33, 2003. 10.1007/s00393-003-0480-4.
- [116] M. Szkudlarek, M. Court-Payen, S. Jacobsen, M. Klarlund, H. S. Thomsen, and M. Ostergaard, "Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 48, no. 4, pp. 955–962, 2003.
- [117] U. Fiocco, L. Cozzi, L. Rubaltelli, C. Rigon, A. De Candia, A. Tregnaghi, C. Gallo, M. A. Favaro, F. Chieco-Bianchi, M. Baldovin, and S. Todesco, "Long-term sonographic follow-up of rheumatoid and psoriatic proliferative knee joint synovitis," *Rheumatology*, vol. 35, no. 2, pp. 155–163, 1996.
- [118] P. J. Lund, A. Heikal, M. J. Maricic, E. A. Krupinski, and C. S. Williams, "Ultrasonographic imaging of the hand and wrist in rheumatoid arthritis," *Skeletal Radiology*, vol. 24, pp. 591–596, 1995. 10.1007/BF00204858.
- [119] R. J. Wakefield, W. W. Gibbon, P. G. Conaghan, P. O'Connor, D. McGonagle, C. Pease, M. J. Green, D. J. Veale, J. D. Isaacs, and P. Emery, "The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: A comparison with conventional radiography," *Arthritis & Rheumatism*, vol. 43, no. 12, pp. 2762–2770, 2000.
- [120] C. S. Winalski, W. E. Palmer, D. I. Rosenthal, and B. N. Weissman, "Magnetic resonance imaging of rheumatoid arthritis," *Radiol. Clin. North Am.*, vol. 34, pp. 243–258, Mar 1996.
- [121] M. Ostergaard, M. Stoltenberg, P. Lovgreen-Nielsen, B. Volck, C. H. Jensen, and I. Lorenzen, "Magnetic resonance imaging-determined synovial membrane and joint effusion volumes in rheumatoid arthritis and osteoarthritis. comparison with the macroscopic and microscopic appearance of the synovium," *Arthritis & Rheumatism*, vol. 40, no. 10, pp. 1856–1867, 1997.

- [122] T. B. Graham, J. S. Blebea, V. Gylys-Moriu, and M. H. Passo, "Magnetic resonance imaging in juvenile rheumatoid arthritis," *Seminars in Arthritis and Rheumatism*, vol. 27, no. 3, pp. 161 – 168, 1997.
- [123] A. L. Tan, R. J. Wakefield, P. G. Conaghan, P. Emery, and D. McGonagle, "Imaging of the musculoskeletal system: magnetic resonance imaging, ultrasonography and computed tomography," *Best Practice & Research Clinical Rheumatology*, vol. 17, no. 3, pp. 513 – 528, 2003. How to Assess Musculoskeletal Conditions.
- [124] M. Backhaus, T. Kamradt, D. Sandrock, D. Loreck, J. Fritz, K. J. Wolf, H. Raber, B. Hamm, G. R. Burmester, and M. Bollow, "Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging," *Arthritis Rheum.*, vol. 42, pp. 1232–1245, Jun 1999.
- [125] I. Eshed, E. Feist, C. E. Althoff, M. Backhaus, G. R. Burmester, B. Hamm, and K.-G. A. Hermann, "Early rheumatoid arthritis-do we really know what it means? Consistency and distribution of MRI findings according to different definitions for early rheumatoid arthritis," *CLINICAL RHEUMATOLOGY*, vol. 30, pp. 551–555, APR 2011.
- [126] J. Narvaez, E. Sirvent, J. A. Narvaez, J. Bas, C. Gomez-Vaquero, D. Reina, J. M. Nolla, and J. Valverde, "Usefulness of magnetic resonance imaging of the hand versus anticyclic citrullinated peptide antibody testing to confirm the diagnosis of clinically suspected early rheumatoid arthritis in the absence of rheumatoid factor and radiographic erosions," *Semin. Arthritis Rheum.*, vol. 38, pp. 101–109, Oct 2008.
- [127] K. Karstila, M. Korpela, S. Sihvonen, and J. Mustonen, "Prognosis of clinical renal disease and incidence of new renal findings in patients with rheumatoid arthritis: follow-up of a population-based study," *Clinical Rheumatology*, vol. 26, pp. 2089–2095, 2007.
- [128] H. J. Helin, M. M. Korpela, J. T. Mustonen, and A. I. Pasternack, "Renal biopsy findings and clinicopathologic correlations in rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 38, no. 2, pp. 242–247, 1995.

- [129] M. A. Perazella, "Current status of gadolinium toxicity in patients with kidney disease," *Clin J Am Soc Nephrol*, vol. 4, pp. 461–469, Feb 2009.
- [130] M. Tamai, A. Kawakami, M. Uetani, S. Takao, H. Rashid, F. Tanaka, K. Fujikawa, T. Aramaki, H. Nakamura, N. Iwanaga, Y. Izumi, K. Arima, K. Aratake, M. Kamachi, M. Huang, T. Origuchi, H. Ida, K. Aoyagi, and K. Eguchi, "Early prediction of rheumatoid arthritis by serological variables and magnetic resonance imaging of the wrists and finger joints: results from prospective clinical examination," *Ann. Rheum. Dis.*, vol. 65, pp. 134–135, Jan 2006.
- [131] V. Prapavat, R. Schuetz, W. Runge, J. Beuthan, and G. J. Mueller, "Evaluation of scattered light distributions of cw-transillumination for functional diagnostic of rheumatic disorders in interphalangeal joints," *Proc. SPIE 2626, Photon Propagation in Tissues*, vol. 2626, no. 1, pp. 121–129, 1995.
- [132] H. Mohamed-Ali, R. W. Hauer, and H. Sorensen, "Morphology and growth behavior of synovial cells in monolayer culture," *Z Rheumatol*, vol. 50, pp. 74–81, 1991.
- [133] L. Dahlberg, L. Ryd, D. Heinegard, and L. S. Lohmander, "Proteoglycan fragments in joint fluid. Influence of arthrosis and inflammation," *Acta Orthop Scand*, vol. 63, pp. 417–423, Aug 1992.
- [134] N. Thumb, G. Kellner, F. Mayrhofer, and A. Wottawa, "Local and systemic parameters in the synovial milieu," *Z Rheumatol*, vol. 44, pp. 20–25, 1985.
- [135] V. Prapavat, W. Runge, J. Mans, A. Krause, J. Beuthan, and G. Muller, "The development of a finger joint phantom for the optical simulation of early inflammatory rheumatic changes," *Biomed Tech (Berl)*, vol. 42, pp. 319–326, Nov 1997.
- [136] V. Prapavat, *Anwendung der experimentellen Systemanalyse zur Informations- Gewinnung aus Streulicht im Frühstadium entzündlich-rheumatischer Veränderungen*. PhD thesis, Technical University Berlin, 1997.
- [137] V. Prapavat, J. Mans, R. Schuetz, G. Regling, J. Beuthan, and G. J. Mueller, "In-vivo investigations on the detection of chronic polyarthritis (c.p.) using a cw-transillumination method at interphalangeal joints," *Proc. SPIE 2626, Photon Propagation in Tissues*, pp. 360–366, 1995.

- [138] A. K. Scheel, A. Krause, I. M.-v. Rheinbaben, G. Metzger, H. Rost, V. Tresp, P. Mayer, M. Reuss-Borst, and G. A. Müller, "Assessment of proximal finger joint inflammation in patients with rheumatoid arthritis, using a novel laser-based imaging technique," *Arthritis & Rheumatism*, vol. 46, no. 5, pp. 1177–1184, 2002.
- [139] A. Schwaighofer, V. Tresp, P. Mayer, A. Krause, J. Beuthan, H. Rost, G. Metzger, G. Muller, and A. Scheel, "Classification of rheumatoid joint inflammation based on laser imaging," *Biomedical Engineering, IEEE Transactions on*, vol. 50, pp. 375–382, march 2003.
- [140] C. D. Klose, A. D. Klose, U. Netz, J. Beuthan, and A. H. Hielscher, "Multiparameter classifications of optical tomographic images," *Journal of Biomedical Optics*, vol. 13, no. 5, p. 050503, 2008.
- [141] L. D. Montejo, J. D. Montejo, H. K. Kim, U. J. Netz, C. D. Klose, S. Blaschke, P. A. Zwaka, G. A. Müller, J. Beuthan, and A. H. Hielscher, "Comparison of classification methods for detection of rheumatoid arthritis with optical tomography," in *Biomedical Optics*, p. BWF2, Optical Society of America, 2010.
- [142] Y. Xu, N. Iftimia, H. Jiang, L. Key, and M. Bolster, "Imaging of in vitro and in vivo bones and joints with continuous-wave diffuse optical tomography," *Opt. Express*, vol. 8, pp. 447–451, Mar 2001.
- [143] Y. Xu, N. Iftimia, H. Jiang, L. L. Key, and M. B. Bolster, "Three-dimensional diffuse optical tomography of bones and joints," *J Biomed Opt*, vol. 7, pp. 88–92, Jan 2002.
- [144] Z. Yuan, Q. Zhang, E. S. Sobel, and H. Jiang, "High-resolution x-ray guided three-dimensional diffuse optical tomography of joint tissues in hand osteoarthritis: morphological and functional assessments," *Med Phys*, vol. 37, pp. 4343–4354, Aug 2010.
- [145] J. Xiao, L. Yao, Y. Sun, E. S. Sobel, J. He, and H. Jiang, "Quantitative two-dimensional photoacoustic tomography of osteoarthritis in the finger joints," *Opt. Express*, vol. 18, pp. 14359–14365, Jul 2010.
- [146] Z. Yuan, Q. Zhang, E. Sobel, and H. Jiang, "Comparison of diffusion approximation and higher order diffusion equations for optical tomography of osteoarthritis," *J Biomed Opt*, vol. 14, p. 054013, 2009.

- [147] X. Wang, D. L. Chamberland, and D. A. Jamadar, “Noninvasive photoacoustic tomography of human peripheral joints toward diagnosis of inflammatory arthritis,” *Opt Lett*, vol. 32, pp. 3002–3004, Oct 2007.
- [148] X. Wang, D. L. Chamberland, P. L. Carson, J. B. Fowlkes, R. O. Bude, D. A. Jamadar, and B. J. Roessler, “Imaging of joints with laser-based photoacoustic tomography: an animal study,” *Med Phys*, vol. 33, pp. 2691–2697, Aug 2006.
- [149] L. L. Gompels, N. H. Lim, T. Vincent, and E. M. Paleolog, “In vivo optical imaging in arthritis—an enlightening future?,” *Rheumatology (Oxford)*, vol. 49, pp. 1436–1446, Aug 2010.
- [150] S. Biswal, D. L. Resnick, J. M. Hoffman, and S. S. Gambhir, “Molecular imaging: integration of molecular imaging into the musculoskeletal imaging practice,” *Radiology*, vol. 244, pp. 651–671, Sep 2007.
- [151] M. H. Brem, P. M. Schlechtweg, J. Mackenzie, C. S. Winalski, and P. Lang, “[Molecular imaging: future uses in arthritides],” *Radiologe*, vol. 46, pp. 394–402, May 2006.
- [152] K. Doi, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198 – 211, 2007. Computer-aided Diagnosis (CAD) and Image-guided Decision Support.
- [153] A. P. G. Leila H. Eadie, Paul Taylor, “A systematic review of computer-assisted diagnosis in diagnostic cancer imaging,” *European Journal of Radiology*, vol. 81, pp. e70–e76, Jan 2012.
- [154] K. Doi, “Current status and future potential of computer-aided diagnosis in medical imaging,” *Br J Radiol*, vol. 78, pp. 3–19, 2005.
- [155] J. Shiraishi, K. Sugimoto, F. Moriyasu, N. Kamiyama, and K. Doi, “Computer-aided diagnosis for the classification of focal liver lesions by use of contrast-enhanced ultrasonography,” *Medical Physics*, vol. 35, no. 5, pp. 1734–1746, 2008.
- [156] L. A. Meinel, A. H. Stolpen, K. S. Berbaum, L. L. Fajardo, and J. M. Reinhardt, “Breast mri lesion classification: Improved performance of human readers with a backpropagation neural network

- computer-aided diagnosis (cad) system,” *Journal of Magnetic Resonance Imaging*, vol. 25, no. 1, pp. 89–95, 2007.
- [157] T. Arazi-Kleinman, P. Causer, R. Jong, K. Hill, and E. Warner, “Can breast mri computer-aided detection (cad) improve radiologist accuracy for lesions detected at mri screening and recommended for biopsy in a high-risk population?,” *Clinical Radiology*, vol. 64, no. 12, pp. 1166 – 1174, 2009.
- [158] K. Suzuki, S. G. A. III, F. Li, S. Sone, and K. Doi, “Massive training artificial neural network (mtann) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography,” *Medical Physics*, vol. 30, no. 7, pp. 1602–1617, 2003.
- [159] S. Katsuragawa and K. Doi, “Computer-aided diagnosis in chest radiography,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 212 – 223, 2007. Computer-aided Diagnosis (CAD) and Image-guided Decision Support.
- [160] T. W. Freer and M. J. Ulissey, “Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center1,” *Radiology*, vol. 220, no. 3, pp. 781–786, 2001.
- [161] K. Drukker, L. Pesce, and M. Giger, “Repeatability in computer-aided diagnosis: Application to breast cancer diagnosis on sonography,” *Medical Physics*, vol. 37, no. 6, pp. 2659–2669, 2010.
- [162] D. Newell, K. Nie, J.-H. Chen, C.-C. Hsu, H. Yu, O. Nalcioğlu, and M.-Y. Su, “Selection of diagnostic features on breast mri to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement,” *European Radiology*, vol. 20, pp. 771–781, 2010. 10.1007/s00330-009-1616-y.
- [163] J. E. Levman, P. Causer, E. Warner, and A. L. Martel, “Effect of the enhancement threshold on the computer-aided detection of breast cancer using mri,” *Academic Radiology*, vol. 16, no. 9, pp. 1064 – 1069, 2009.
- [164] W. B. DeMartini, C. D. Lehman, S. Peacock, and M. T. Russell, “Computer-aided detection applied to breast mri: Assessment of cad-generated enhancement and tumor sizes in breast cancers before and after neoadjuvant chemotherapy,” *Academic Radiology*, vol. 12, no. 7, pp. 806 – 814, 2005.

- [165] X. Qi, J. Michael V. Sivak, G. Isenberg, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in barrett's esophagus using endoscopic optical coherence tomography," *Journal of Biomedical Optics*, vol. 11, no. 4, p. 044010, 2006.
- [166] F. Bazant-Hegemark and N. Stone, "Towards automated classification of clinical optical coherence tomography data of dense tissues," *Lasers in Medical Science*, vol. 24, pp. 627–638, 2009. 10.1007/s10103-008-0615-6.
- [167] X. Song, B. W. Pogue, S. Jiang, M. M. Doyley, H. Dehghani, T. D. Tosteson, and K. D. Paulsen, "Automated region detection based on the contrast-to-noise ratio in near-infrared tomography," *Appl. Opt.*, vol. 43, pp. 1053–1062, Feb 2004.
- [168] X. Song, B. W. Pogue, H. Dehghani, S. Jiang, K. D. Paulsen, and T. D. Tosteson, "Receiver operating characteristic and location analysis of simulated near-infrared tomography images," *J Biomed Opt*, vol. 12, no. 5, p. 054013, 2007.
- [169] C. Zhu, T. M. Breslin, J. Harter, and N. Ramanujam, "Model based and empirical spectral analysis for the diagnosis of breast cancer," *Opt. Express*, vol. 16, pp. 14961–14978, Sep 2008.
- [170] T. M. Bydlon, S. A. Kennedy, L. M. Richards, J. Q. Brown, B. Yu, M. K. Junker, J. Gallagher, J. Geradts, L. G. Wilke, and N. Ramanujam, "Performance metrics of an optical spectral imaging system for intra-operative assessment of breast tumor margins," *Opt Express*, vol. 18, pp. 8058–8076, Apr 2010.
- [171] D. R. Busch, W. Guo, R. Choe, T. Durduran, M. D. Feldman, C. Mies, M. A. Rosen, M. D. Schnall, B. J. Czerniecki, J. Tchou, A. DeMichele, M. E. Putt, and A. G. Yodh, "Computer aided automatic detection of malignant lesions in diffuse optical mammography," *Medical Physics*, vol. 37, no. 4, pp. 1840–1849, 2010.
- [172] D. R. Busch, R. Choe, M. A. Rosen, W. Guo, T. Durduran, M. D. Feldman, C. Mies, B. J. Czerniecki, J. Tchou, A. Demichele, M. D. Schnall, and A. G. Yodh, "Optical malignancy parameters for monitoring progression of breast cancer neoadjuvant chemotherapy," *Biomed Opt Express*, vol. 4, pp. 105–121, Jan 2013.

- [173] C. D. Klose, A. D. Klose, U. J. Netz, A. K. Scheel, J. Beuthan, and A. H. Hielscher, "Computer-aided interpretation approach for optical tomographic images," *Journal of Biomedical Optics*, vol. 15, no. 6, p. 066020, 2010.
- [174] O. Minet, U. Zabarylo, and J. Beuthan, "Deconvolution of laser based images for monitoring rheumatoid arthritis," *Laser Physics Letters*, vol. 2, pp. 556–563, 2005.
- [175] K. Ren, G. Bal, and A. H. Hielscher, "Frequency domain optical tomography based on the equation of radiative transfer," *SIAM J. Sci. Comput.*, vol. 28, pp. 1463–1489, Apr. 2006.
- [176] D. M. van der Heijde, M. A. van 't Hof, P. L. van Riel, L. A. Theunisse, E. W. Lubberts, M. A. van Leeuwen, M. H. van Rijswijk, and L. B. van de Putte, "Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score," *Ann. Rheum. Dis.*, vol. 49, pp. 916–920, Nov 1990.
- [177] W. M. Wraight, J. D. Smith, and D. Floyd, "Nomenclature for fingers and phalanges: to name or to number?," *J Plast Reconstr Aesthet Surg*, vol. 60, no. 4, pp. 368–371, 2007.
- [178] M. Ostergaard, C. Peterfy, P. Conaghan, F. McQueen, P. Bird, B. Ejbjerg, R. Shnier, P. O'Connor, M. Klarlund, P. Emery, H. Genant, M. Lassere, and J. Edmonds, "OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system," *J. Rheumatol.*, vol. 30, pp. 1385–1386, Jun 2003.
- [179] C. Schirmer, A. K. Scheel, C. E. Althoff, T. Schink, I. Eshed, A. Lembcke, G. R. Burmester, M. Backhaus, B. Hamm, and K. G. Hermann, "Diagnostic quality and scoring of synovitis, tenosynovitis and erosions in low-field MRI of patients with rheumatoid arthritis: a comparison with conventional MRI," *Ann. Rheum. Dis.*, vol. 66, pp. 522–529, Apr 2007.
- [180] H. A. Fuchs, R. H. Brooks, L. F. Callahan, and T. Pincus, "A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis," *Arthritis Rheum.*, vol. 32, pp. 531–537, May 1989.
- [181] M. L. Prevoo, M. A. van 't Hof, H. H. Kuper, M. A. van Leeuwen, L. B. van de Putte, and P. L. van Riel, "Modified disease activity scores that include twenty-eight-joint counts. Development and

- validation in a prospective longitudinal study of patients with rheumatoid arthritis,” *Arthritis Rheum.*, vol. 38, pp. 44–48, January 1995.
- [182] U. J. Netz, J. Beuthan, and A. H. Hielscher, “Multipixel system for gigahertz frequency-domain optical imaging of finger joints,” *Review of Scientific Instruments*, vol. 79, no. 3, p. 034301, 2008.
- [183] L. C. Henyey and J. L. Greenstein, “Diffuse radiation in the galaxy,” *Astrophysical Journal*, vol. 93, pp. 70–83, 1941.
- [184] M. Modest, *Radiative Heat Transfer*. Chemical, Petrochemical & Process, ACADEMIC PressINC, 2003.
- [185] M. H. Zweig and G. Campbell, “Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine,” *Clin. Chem.*, vol. 39, pp. 561–577, Apr 1993.
- [186] S. J. Dixon and R. G. Brereton, “Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure,” *Chemometrics and Intelligent Laboratory Systems*, vol. 95, no. 1, pp. 1 – 17, 2009.
- [187] G. McLachlan, *Discriminant analysis and statistical pattern recognition*. Wiley series in probability and mathematical statistics, Wiley, 1992.
- [188] H. K. Kim, U. J. Netz, J. Beuthan, and A. H. Hielscher, “Optimal source-modulation frequencies for transport-theory-based optical tomography of small-tissue volumes,” *Opt Express*, vol. 16, pp. 18082–18101, Oct 2008.
- [189] S. Theodoridis and K. Koutroubas, *Pattern recognition*. Elsevier/Academic Press, 2006.
- [190] S. Glantz, *Primer of biostatistics*. No. pp. 1-72 in *Primer of Biostatistics*, McGraw-Hill Medical Pub. Division, 2005.
- [191] S. Killip, Z. Mahfoud, and K. Pearce, “What is an intraclass correlation coefficient? crucial concepts for primary care researchers,” *The Annals of Family Medicine*, vol. 2, no. 3, pp. 204–208, 2004.

- [192] A. Donner and J. J. Koval, "The estimation of intraclass correlation in the analysis of family data," *Biometrics*, vol. 36, no. 1, pp. 19–25, 1980.
- [193] E. B. Wilson, "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.
- [194] H. Ahrens, "Multivariate variance-covariance components (mvcc) and generalized intraclass correlation coefficient (gicc)," *Biometrische Zeitschrift*, vol. 18, no. 7, pp. 527–533, 1976.
- [195] C. Bishop, *Pattern recognition and machine learning*. Information science and statistics, Springer, 2006.
- [196] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines: and other kernel-based learning methods*. Cambridge University Press, 2006.
- [197] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [198] T. Bäck, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, 1996.
- [199] Z. Szekanecz, T. Besenyei, A. Szentpétery, and A. E. Koch, "Angiogenesis and vasculogenesis in rheumatoid arthritis," *Curr Opin Rheumatol*, vol. 22, pp. 299–306, May 2010.
- [200] S. Karie, F. Gandjbakhch, N. Janus, V. Launay-Vacher, S. Rozenberg, C. U. Mai Ba, P. Bourgeois, and G. Deray, "Kidney disease in RA patients: prevalence and implication on RA-related drugs management: the MATRIX study," *Rheumatology (Oxford)*, vol. 47, pp. 350–354, Mar 2008.
- [201] J. E. Freeston, R. J. Wakefield, P. G. Conaghan, E. M. A. Hensor, S. P. Stewart, and P. Emery, "A diagnostic algorithm for persistence of very early inflammatory arthritis: the utility of power doppler ultrasound when added to conventional assessment tools," *Annals of the Rheumatic Diseases*, vol. 69, no. 2, pp. 417–419, 2010.
- [202] W. Minkowycz, E. Sparrow, and J. Murthy, *Handbook of Numerical Heat Transfer*. Wiley, 2006.

- [203] Y. Saad and M. H. Schultz, “Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems,” *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 3, pp. 856–869, 1986.
- [204] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, second ed., 2003.
- [205] J. Nocedal and S. Wright, *Numerical optimization*. Springer Series in Operations Research and Financial Engineering, Springer Science+Business Media, LLC., 2006.
- [206] K. E. Atkinson, *An Introduction to Numerical Analysis*. John Wiley & Sons, Canada, 1989.
- [207] D. Boas, C. Pitris, and N. Ramanujam, *Handbook of Biomedical Optics*. Taylor & Francis, 2011.
- [208] M. S. Patterson and B. W. Pogue, “Mathematical model for time-resolved and frequency-domain fluorescence spectroscopy in biological tissues,” *Appl. Opt.*, vol. 33, pp. 1963–1974, Apr 1994.
- [209] J. M. Lasker, J. M. Masciotti, M. Schoenecker, C. H. Schmitz, and A. H. Hielscher, “Digital-signal-processor-based dynamic imaging system for optical tomography,” *Review of Scientific Instruments*, vol. 78, no. 8, p. 083706, 2007.
- [210] B. W. Pogue and G. Burke, “Fiber-optic bundle design for quantitative fluorescence measurement from tissue,” *Appl. Opt.*, vol. 37, pp. 7429–7436, Nov 1998.
- [211] R. B. Simpson, “Automatic local refinement for irregular rectangular meshes,” *International Journal for Numerical Methods in Engineering*, vol. 14, no. 11, pp. 1665–1678, 1979.
- [212] M. Berger and P. Colella, “Local adaptive mesh refinement for shock hydrodynamics,” *Journal of Computational Physics*, vol. 82, no. 1, pp. 64 – 84, 1989.
- [213] W. L. Chen, F. S. Lien, and M. A. Leschziner, “Local mesh refinement within a multi-block structured-grid scheme for general flows,” *Computer Methods in Applied Mechanics and Engineering*, vol. 144, no. 3-4, pp. 327 – 369, 1997.
- [214] M. J. Berger and J. Oliger, “Adaptive mesh refinement for hyperbolic partial differential equations,” *Journal of Computational Physics*, vol. 53, no. 3, pp. 484 – 512, 1984.

- [215] J. P. Jessee, W. A. Fiveland, L. H. Howell, P. Colella, and R. B. Pember, “An adaptive mesh refinement algorithm for the radiative transport equation,” *Journal of Computational Physics*, vol. 139, no. 2, pp. 380 – 398, 1998.
- [216] A. Joshi, W. Bangerth, K. Hwang, J. C. Rasmussen, and E. M. Sevick-Muraca, “Fully adaptive fem based fluorescence optical tomography from time-dependent measurements with area illumination and detection,” *Medical Physics*, vol. 33, no. 5, pp. 1299–1310, 2006.
- [217] A. H. Hielscher, “Optical tomographic imaging of small animals,” *Current Opinion in Biotechnology*, vol. 16, no. 1, pp. 79 – 88, 2005. Analytical biotechnology.
- [218] A. Soubret, J. Ripoll, and V. Ntziachristos, “Accuracy of fluorescent tomography in the presence of heterogeneities: study of the normalized born ratio,” *Medical Imaging, IEEE Transactions on*, vol. 24, pp. 1377 –1386, Oct. 2005.

Forward Model

A.1 Conversion of FEM mesh into FVM mesh

The pseudo code for converting the 3D GID finite element mesh to finite volume mesh.

Algorithm 1: Conversion of FEM mesh into FVM mesh

Pseudo code:

Step 1: Go over all the tetrahedron elements. For each tetrahedron \mathbf{E}_{i_E} with the nodes $(\mathbf{p}_{i_p}, \mathbf{p}_{j_p}, \mathbf{p}_{k_p}, \mathbf{p}_{l_p})$, we record

- The coordinates of the centroid of \mathbf{E}_{i_E} : $\mathbf{c}_{i_E}^E = (\mathbf{p}_{i_p} + \mathbf{p}_{j_p} + \mathbf{p}_{k_p} + \mathbf{p}_{l_p})/4$.
- All the triangle faces of the tetrahedron element \mathbf{E}_{i_E} .

Step 2: Go over all the triangle faces after removing those duplicate ones. For each face \mathbf{F}_{i_F} with nodes $(\mathbf{p}_{i_p}, \mathbf{p}_{j_p}, \mathbf{p}_{k_p})$, we record

- The indices of tetrahedron elements that \mathbf{F}_{i_F} belongs to. (Practically these were record in step 1)
- The coordinates of the centroid of \mathbf{F}_{i_F} : $\mathbf{c}_{i_F}^F = (\mathbf{p}_{i_p} + \mathbf{p}_{j_p} + \mathbf{p}_{k_p})/3$.
- All the edges of the face \mathbf{F}_{i_F} .

step 3: Go over all the edges after removing those duplicate ones. For each edge \mathbf{e}_{i_e} , with nodes $(\mathbf{p}_{i_p}, \mathbf{p}_{j_p})$, we record

- The indices of triangle faces that \mathbf{e}_{i_e} belongs to. (Practically these were record in step 2)
- The coordinates of the centroid of \mathbf{e}_{i_e} : $\mathbf{c}_{i_e}^e = (\mathbf{p}_{i_p} + \mathbf{p}_{j_p})/2$.
- The node indices i_p and j_p .

Step 4: Go over all the nodes. For each node \mathbf{p}_{i_p} , we record

- The indices of edges that \mathbf{p}_{i_p} belongs to. (Practically these were record in step 3)

Step 5: Build the finite volume mesh and gather the information as below:

for every node \mathbf{p}_{i_p}

Construct a finite volume element \mathbf{V}_{i_p} centered at \mathbf{p}_{i_p} ;

Set the volume of this element $V_{i_p} = 0$;

Set the face number of this element $n_{i_p} = 0$;

for every edge \mathbf{e}_{i_e} that \mathbf{p}_{i_p} belongs to

Let \mathbf{p}_{j_p} be the other node that belongs to \mathbf{e}_{i_e} ;

Let $\mathbf{n} = \mathbf{p}_{j_p} - \mathbf{p}_{i_p}$;

for every face \mathbf{F}_{i_F} that \mathbf{e}_{i_e} belongs to

for every element \mathbf{E}_{i_E} that \mathbf{F}_{i_F} belongs to

Add a face formed by $(\mathbf{c}_{i_e}^e, \mathbf{c}_{i_F}^F, \mathbf{c}_{i_E}^E)$ to \mathbf{V}_{i_p} ;

The outgoing normal vector of this face is \mathbf{n} ;

The neighbor element to \mathbf{V}_{i_p} on this face is \mathbf{V}_{j_p} ;

Let ΔV be the volume of tetrahedron $(\mathbf{p}_{i_p}, \mathbf{c}_{i_e}^e, \mathbf{c}_{i_F}^F, \mathbf{c}_{i_E}^E)$;

$V_{i_p} = V_{i_p} + \Delta V$;

$n_{i_p} = n_{i_p} + 1$;

end for

if \mathbf{F}_{i_F} is on the boundary (it only belongs to one element)

Add a face formed by $(\mathbf{p}_{i_p}, \mathbf{c}_{i_e}^e, \mathbf{c}_{i_F}^F)$ to \mathbf{V}_{i_p} ;

Compute the outgoing normal vector for this face;

The neighbor element to \mathbf{V}_{i_p} on this face is *Null*;

$n_{i_p} = n_{i_p} + 1$;

end if
end for
end for
end for



The pseudo code for converting the 2D GID finite element mesh to finite volume mesh.



Pseudo code:

Step 1: Go over all the triangle faces. For each face \mathbf{F}_{i_F} with nodes $(\mathbf{p}_{i_p}, \mathbf{p}_{j_p}, \mathbf{p}_{k_p})$, we record

- The coordinates of the centroid of \mathbf{F}_{i_F} : $\mathbf{c}_{i_F}^F = (\mathbf{p}_{i_p} + \mathbf{p}_{j_p} + \mathbf{p}_{k_p})/3$.
- All the edges of the face \mathbf{F}_{i_F} .

step 2: Go over all the edges after removing those duplicate ones. For each edge \mathbf{e}_{i_e} , with nodes $(\mathbf{p}_{i_p}, \mathbf{p}_{j_p})$, we record

- The indices of triangle faces that \mathbf{e}_{i_e} belongs to. (Practically these were record in step 2)
- The coordinates of the centroid of \mathbf{e}_{i_e} : $\mathbf{c}_{i_e}^e = (\mathbf{p}_{i_p} + \mathbf{p}_{j_p})/2$.
- The node indices i_p and j_p .

Step 3: Go over all the nodes. For each node \mathbf{p}_{i_p} , we record

- The indices of edges that \mathbf{p}_{i_p} belongs to. (Practically these were record in step 2)

Step 4: Build the finite volume mesh and gather the information as below:

for every node \mathbf{p}_{i_p}

Construct a finite volume element \mathbf{V}_{i_p} centered at \mathbf{p}_{i_p} ;

Set the area of this element $V_{i_p} = 0$;

Set the edge number of this element $n_{i_p} = 0$;

for every edge \mathbf{e}_{i_e} that \mathbf{p}_{i_p} belongs to

Let \mathbf{p}_{j_p} be the other node that belongs to \mathbf{e}_{i_e} ;

Let $\mathbf{n} = \mathbf{p}_{j_p} - \mathbf{p}_{i_p}$;

for every face \mathbf{F}_{i_F} that \mathbf{e}_{i_e} belongs to

Add a edge formed by $(\mathbf{c}_{i_e}^e, \mathbf{c}_{i_F}^F)$ to \mathbf{V}_{i_p} ;

The outgoing normal vector of this edge is \mathbf{n} ;

The neighbor element to \mathbf{V}_{i_p} on this edge is \mathbf{V}_{j_p} ;

Let ΔV be the area of triangle $(\mathbf{p}_{i_p}, \mathbf{c}_{i_e}^e, \mathbf{c}_{i_F}^F)$;

$$V_{i_p} = V_{i_p} + \Delta V;$$

$$n_{i_p} = n_{i_p} + 1;$$

end for

if \mathbf{e}_{i_e} is on the boundary (it only belongs to one face)

Add a edge formed by $(\mathbf{p}_{i_p}, \mathbf{c}_{i_e}^e)$ to \mathbf{V}_{i_p} ;

Compute the outgoing normal vector for this edge;

The neighbor element to \mathbf{V}_{i_p} on this edge is *Null*;

$$n_{i_p} = n_{i_p} + 1;$$

end if

end for

end for



Inverse Problem

B.1 L-BFGS algorithm

The L-BFGS algorithm is defined by Algorithms 8 and 9.

Algorithm 8 L-BFGS (outer loop) [205]

```

Choose starting point  $x_0$ , integer  $m > 0$ 
 $k \leftarrow 0$ 
while rSQP do
    Choose  $H_k^0 \left( H_k^0 = \gamma_k I, \gamma_k = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}} \right)$ 
    Compute  $p_k \leftarrow -H_k \nabla f_k$  (See inner loop details in Algorithm 9)
    Compute  $x_{k+1} \leftarrow x_k + \alpha_k p_k$ 
    if  $k > m$  then
        Discard the vector pair  $\{s_{k-m}, y_{k-m}\}$  from storage
    end if
    Compute and save  $s_k \leftarrow x_{k+1} - x_k, y_k = \nabla f_{k+1} - \nabla f_k$ 
     $k \leftarrow k + 1$ 
end while

```

Algorithm 9 L-BFGS (inner loop) [205]

```
q ← ∇f
for i = k - 1, k - 2, ..., k - m do
    ρi =  $\frac{1}{y_i^\top s_i}$ 
    αi = ρi si⊤ q
    q = q - αi yi
end for
r ← Hk0 q
for i = k - m, k - m + 1, ..., k - 1 do
    β ← ρi yi⊤ r
    r ← si (αi - β)
end for
∴ Hk ∇fk = r
```

B.2 Right hand side of adjoint equation

The derivation of the terms associated with the right-hand side of the adjoint equation are derived in this section. In particular, we are interested in the gradient of the objective function, u , with respect to the forward variable, u . It is important, however, to note that u is a complex-valued vector and, therefore, care must be taken when performing the necessary differentiations. First, expand the objective function into its real and imaginary parts.

$$\begin{aligned} f(\mu, u) &= \frac{1}{2} (Qu - M)^\top \overline{(Qu - M)} \\ &= \frac{1}{2} [(Qu_{\text{Re}} - M_{\text{Re}}) + i(Qu_{\text{Im}} - M_{\text{Im}})]^\top [(\overline{Qu_{\text{Re}} - M_{\text{Re}}}) + i(\overline{Qu_{\text{Im}} - M_{\text{Im}}})] \\ &= \frac{1}{2} [(Qu_{\text{Re}} - M_{\text{Re}}) + i(Qu_{\text{Im}} - M_{\text{Im}})]^\top [(Qu_{\text{Re}} - M_{\text{Re}}) - i(Qu_{\text{Im}} - M_{\text{Im}})] \\ &= \frac{1}{2} \left[(Qu_{\text{Re}} - M_{\text{Re}})^\top (Qu_{\text{Re}} - M_{\text{Re}}) - (Qu_{\text{Im}} - M_{\text{Im}})^\top (Qu_{\text{Im}} - M_{\text{Im}}) \right] \\ &= \frac{1}{2} \left[\left(u_{\text{Re}}^\top Q^\top Qu_{\text{Re}} - M_{\text{Re}}^\top Qu_{\text{Re}} + M_{\text{Re}}^\top M_{\text{Re}} \right) + \left(u_{\text{Im}}^\top Q^\top Qu_{\text{Im}} - M_{\text{Im}}^\top Qu_{\text{Im}} + M_{\text{Im}}^\top M_{\text{Im}} \right) \right]. \end{aligned} \tag{B.1}$$

The partial derivative of f with respect to the complex-valued forward variable u can be decomposed into a real and imaginary part,

$$\frac{\partial f(\mu, u)}{\partial u} = \frac{\partial f(\mu, u)}{\partial u_{\text{Re}}} + \text{i} \frac{\partial f(\mu, u)}{\partial u_{\text{Im}}}. \quad (\text{B.2})$$

Then, differentiating (B.1) according to (B.2) yields the sought after expression,

$$\begin{aligned} \frac{\partial f(\mu, u)}{\partial u} &= Q^\top (Qu_{\text{Re}} - M_{\text{Re}}) - \text{i} Q^\top (Qu_{\text{Im}} - M_{\text{Im}}) \\ &= Q^\top [(Qu_{\text{Re}} - \text{i} Qu_{\text{Im}}) - (M_{\text{Re}} - \text{i} M_{\text{Im}})] \\ &= Q^\top \overline{(Qu - M)}. \end{aligned} \quad (\text{B.3})$$

B.3 Partial derivative of u

The partial derivate of the the forward variable $u(\mu)$ with respect to the inverse variable is needed in order to compute the gradient of the objective function using the adjoint equation method. The computation of this quantity is as follows.

$$\frac{\partial}{\partial \mu} A(\mu) u(\mu) = u(\mu)^\top \frac{\partial A(\mu)}{\partial \mu} + A(\mu) \frac{\partial u(\mu)}{\partial \mu} \quad (\text{B.4})$$

Setting (B.4) equal to zero allows for the following simplification.

$$\frac{\partial u(\mu)}{\partial \mu} = -A(\mu)^{-1} \left(u(\mu)^\top \frac{\partial A(\mu)}{\partial \mu} \right) \quad (\text{B.5})$$

Alternatively,

$$\frac{\partial u(\mu)}{\partial \mu} = -\frac{\partial A u}{\partial \mu} \left(A^\top \right)^{-1}. \quad (\text{B.6})$$

B.4 Gradient of objective function

The expression for the gradient of the objective function of the unconstrained problem is

$$\nabla_{\mu} f(\mu, u) = -\lambda^{\top} \frac{\partial Au}{\partial \mu}, \quad (\text{B.7})$$

where it was shown in (9.21) that λ , the adjoint variable, is defined as

$$\lambda = A^{-1} Q^{\top} \overline{(Qu - M)}. \quad (\text{B.8})$$

The final version of the gradient is found by decomposing the forward problem and the objective function into an equivalent system that explicitly accounts for the real and imaginary parts as follows.

$$Au = b$$

$$\begin{bmatrix} A_{\text{Re}} & -A_{\text{Im}} \\ A_{\text{Im}} & A_{\text{Re}} \end{bmatrix} \begin{bmatrix} u_{\text{Re}} \\ u_{\text{Im}} \end{bmatrix} = \begin{bmatrix} b_{\text{Re}} \\ b_{\text{Im}} \end{bmatrix}$$

Then,

$$\begin{aligned}
\lambda^\top \frac{\partial A u}{\partial \mu} &= \begin{bmatrix} \lambda_{\text{Re}} \\ \lambda_{\text{Im}} \end{bmatrix}^\top \frac{\partial}{\partial \mu} \left(\begin{bmatrix} A_{\text{Re}} & -A_{\text{Im}} \\ A_{\text{Im}} & A_{\text{Re}} \end{bmatrix} \begin{bmatrix} u_{\text{Re}} \\ u_{\text{Im}} \end{bmatrix} \right) \\
&= \begin{bmatrix} \lambda_{\text{Re}} \\ \lambda_{\text{Im}} \end{bmatrix}^\top \frac{\partial}{\partial \mu} \begin{bmatrix} A_{\text{Re}} u_{\text{Re}} - A_{\text{Im}} u_{\text{Im}} \\ A_{\text{Im}} u_{\text{Re}} + A_{\text{Re}} u_{\text{Im}} \end{bmatrix} \\
&= \begin{bmatrix} \lambda_{\text{Re}} \\ \lambda_{\text{Im}} \end{bmatrix}^\top \frac{\partial}{\partial \mu} \begin{bmatrix} A_{\text{Re}} u_{\text{Re}} \\ A_{\text{Re}} u_{\text{Im}} \end{bmatrix} \\
&= u_{\text{Re}}^\top \frac{\partial A_{\text{Re}}}{\partial \mu} \lambda_{\text{Re}} + u_{\text{Im}}^\top \frac{\partial A_{\text{Im}}}{\partial \mu} \lambda_{\text{Im}} \\
&= \left(u^\top \frac{\partial A}{\partial \mu} \lambda \right)_{\text{Re}}.
\end{aligned}$$

The final expression for the gradient of the objective function is

$$\nabla_\mu f(\mu, u) = - \left(u^\top \frac{\partial A}{\partial \mu} \lambda \right)_{\text{Re}}. \quad (\text{B.9})$$

B.5 Reduced gradient for rSQP method

The reduced gradient (9.44) is given by

$$g_r^k = Y^{k\top} C^k. \quad (\text{B.10})$$

Then, using the definitions of $Y^{k\top}$ and C^k , the following simplification steps can be taken:

$$\begin{aligned}
g_r^k &= Y^{k\top} C^k \\
&= \mathbf{I} C_\mu^k + \hat{Y}^{k\top} C_u^k \\
&= \hat{Y}^{k\top} Q^\top (Qu^k - M) \\
&= - \left(\begin{pmatrix} (f_u^k)^{-1} & f_\mu^k \end{pmatrix}^\top \overline{(Qu^k - M)} \right) \\
&= - \left(A^{-1} \frac{\partial (A^k u^k)}{\partial \mu} \right)^\top \overline{(Qu^k - M)} \\
&= - \left(\frac{\partial (A^k u^k)}{\partial \mu} \right)^\top (A^\top)^{-1} \overline{(Qu^k - M)} \\
&= - \left[(A^{k-1} \overline{(Qu^k - M)}) \right]^\top \left(\frac{\partial (A^k u^k)}{\partial \mu} \right) \\
&= - \lambda^{k\top} \left(\frac{\partial (A^k u^k)}{\partial \mu} \right) \\
&= - \left(u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k \right)_{\text{Re}}.
\end{aligned} \tag{B.11}$$

The derivation of the last step is presented in detail in Appendix B.4. Note that the following algebraic relationships are used: $A^\top = A$, $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top \mathbf{A}^\top$, $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$, and $\frac{\partial}{\partial \mu}(Au) = u^\top \frac{\partial}{\partial \mu} A$.

B.6 Update direction for the forward variable

Derivation of (9.46) and (9.55) from the second equation in (9.34) is shown below. The quantity Δu^k is the direction along which the forward variable, u , is updated via Newton's method.

$$\begin{aligned}
 \Delta u^k &= -(f_u^k)^{-1} f^k - (f_u^k)^{-1} f_\mu^k \Delta \mu^k \\
 &= -(f_u^k)^{-1} f^k + (f_u^k)^{-1} f_\mu^k (H_r^k)^{-1} (g_r^k) \\
 &= \hat{Z}^k - \hat{Y}^k (H_r^k)^{-1} (g_r^k) \\
 &= -\left(A^k\right)^{-1} (Au - b)^k + \left(A^k\right)^{-1} \left(u^\top \frac{\partial}{\partial \mu} A\right)^k \left(H_r^k\right)^{-1} (g_r^k) \\
 &= -\left(A^k\right)^{-1} \left[(Au - b)^k - \left(u^\top \frac{\partial}{\partial \mu} A\right)^k \left(H_r^k\right)^{-1} (g_r^k) \right] \\
 &= -\left(A^k\right)^{-1} \left[(Au - b)^k + \left(u^\top \frac{\partial}{\partial \mu} A\right)^k \Delta \mu^k \right]
 \end{aligned} \tag{B.12}$$

B.7 Merit function

The merit function used with the PDE-constrained algorithm, as detailed in Section 9.3.3, is the following

$$\phi(\mu^k, u^k) = \frac{1}{2} \left(Qu^k - M \right)^\top \overline{(Qu^k - M)} + \eta_k \left\| A^k u^k - b \right\|_1 \quad (\text{B.13})$$

The directional derivative of $\phi_{\eta_k}(\mu^k, u^k)$, a necessary quantity for the line search subroutine, is obtained as follows.

$$\begin{aligned} D\phi(\mu^k, u^k) &= \nabla_\mu \phi(\mu^k, u^k) \cdot \Delta\mu^k + \nabla_u \phi(\mu^k, u^k) \cdot \Delta u^k - \eta_k \left\| A^k u^k - b \right\|_1 \\ &= \nabla_\mu \phi(\mu^k, u^k) \cdot \Delta\mu^k + \nabla_u \phi(\mu^k, u^k) \cdot \Delta u^k - \eta_k \left\| A^k u^k - b \right\|_1 \\ &= \nabla_\mu \phi(\mu^k, u^k) \cdot \Delta\mu^k - \eta_k \left\| A^k u^k - b \right\|_1 \\ &= \left(u^{k\top} \frac{\partial A^k}{\partial \mu} \lambda^k \right)_{\text{Re}}^\top \Delta\mu^k - \eta_k \left\| A^k u^k - b \right\|_1 \end{aligned} \quad (\text{B.14})$$

B.8 Right hand side of adjoint equation: clinical data

The objective function that is used when considering clinical data generated with unknown source power is presented in Section 10.1 and reproduced below:

$$f(\mu, u) = \frac{1}{2} \frac{1}{M^\top M} \left(M - \frac{Qu}{\tilde{p}} \right)^\top \overline{\left(M - \frac{Qu}{\tilde{p}} \right)}. \quad (\text{B.15})$$

The average intensity (or average predicted detector reading) is a complex valued number and defined as,

$$\tilde{p} = \frac{1}{SD} (Qu)^\top \mathbf{1}. \quad (\text{B.16})$$

Where $\mathbf{1}$ is a vector of ones, S is the total number of sources, and D is the total number of detector per source. Here we assume each source has the same number of detector; however, the above equation can be

easily modified to account for different source-detector configurations. The right hand side of the adjoint equation can be constructed by differentiating f with respect to the forward variable, u , as follows.

$$\begin{aligned}
\frac{\partial}{\partial u} f(\mu, u) &= \frac{\partial}{\partial u} \frac{1}{2} \left(M - \frac{Qu}{\tilde{p}} \right)^\top \overline{\left(M - \frac{Qu}{\tilde{p}} \right)} \\
&= -\frac{\partial}{\partial u} \left[\frac{1}{2} \frac{(Qu)^\top Qu}{\tilde{p}^2} - \frac{M^\top Qu}{\tilde{p}} + \frac{1}{2} M^\top M \right] \\
&= -\frac{Q^\top (Qu)}{\tilde{p}^2} - \frac{(Qu)^\top Qu \cdot \frac{\partial \tilde{p}}{\partial u}}{\tilde{p}^3} - \frac{Q^\top M}{\tilde{p}} + \frac{M^\top Qu \cdot \frac{\partial \tilde{p}}{\partial u}}{\tilde{p}^2} \\
&= -\frac{Q^\top (P - M)}{\tilde{p}} - \frac{P^\top (P - M)}{\tilde{p}} \frac{\partial \tilde{p}}{\partial u} \\
&= -Q^\top \left[\frac{(P - M)^* - P^\top (P - M)^* \frac{1}{SD} \mathbf{1}}{\tilde{p}} \right]
\end{aligned} \tag{B.17}$$

Where the following terms was used,

$$\frac{\partial \tilde{p}}{\partial u} = Q^\top \frac{1}{SD} \mathbf{1}, \tag{B.18}$$

Block-Structured Grids

A deterrent to more widespread implementation of the ERT on structured grids is computational cost. Finely discretized grids are often required when modeling light transport on complex geometries, as it is often important for the numerical grid to match the physical boundary. It is important to note that this can only be achieved with finely discretized grids. However, solving the ERT on a finely discretized grid can be prohibitively expensive. Performing calculations on a coarse grid to save computation time can lead to significant error, which is further compounded if partly reflective boundary conditions are considered.

In addition, a grid that is too coarse at the boundary might not be able to accurately locate boundary sources and detectors. Fiber bundles are typically used to direct photons from a laser source to the medium. Similar sets of fibers are used to direct the escaping photons to a detector [208, 209, 210]. Thus, these sources and detectors are highly localized and a finely discretized grid is necessary to accurately describe their position on the surface. Additionally, in the area of small animal fluorescence and bioluminescence imaging, where it is possible to have internal sources close to the tissue boundary, it is well known that this types of light source can lead to severe modeling errors [67]. Thus, a coarse grid can potentially fail to accurately model the boundary effects and this can lead to further error. For this reason, increased grid refinement near the boundary is desired, and it is something that has not been explored to date for the ERT equations.

These problems can be overcome by using an algorithm that solves the ERT on a grid that is refined only near the object boundary. This results in a grid with various levels of refinement; the grid is coarsely discretized in the interior and finely discretized near the boundary. These types of grids are called block-

structured grids (BSGs) and are a subset of the more general topic known as “adaptive mesh refinement” [75]. A single dense Cartesian grid is transformed into a relatively sparse grid. Thus, solving the ERT on a BSG requires less computational effort than solving it on a single dense grid as there are less grid nodes inside the computational domain.

The subject of generating Cartesian grids with local refinements has itself been a topic of research [75,211]. BSGs have been particularly studied in the field of computational fluid dynamics, especially in the study of shock-hydrodynamics [212] and general fluid flows [213]. Previous work has mainly focused on solving numerical models using the finite volume method (FVM) on BSGs [214].

In connection with the ERT, only one group has used BSGs to date. In 1998, Jesse et al. [215] presented a method to solve the time-independent (or steady-state) ERT on BSGs using the an FVM discretization technique. Their code was limited to two-dimensional rectangular media with isotropic scattering and did not consider partially reflective boundary conditions. While useful for applications in nuclear physics and heat transfer, this algorithm is not suited for problems concerning light propagation in arbitrarily shaped media with highly anisotropic scattering and partially reflective boundary conditions; thus, the algorithm has limited utility in modeling light propagation in biological tissue. In the area of tissue optics, BSGs have been employed for solving the diffusion equation for fluorescent light propagation [216]. That work was limited to treating rectangular geometries and was developed as the finite-element method.

The work presented in this chapter goes beyond the approaches previously presented and implement the first frequency-domain (FD) ERT with reflective boundary conditions on BSGs of arbitrary shape. The FD-ERT is particularly important in small animal imaging, where the diffusion approximation has limited validity, since light travels only a few mean-free path and non-diffusive boundary conditions dominate the solutions [58, 77, 217]. It is highly desirable to have a fast and accurate numerical solver for this and other similar applications (e.g. imaging of arthritic human finger joints [13]) that require solutions to the problem of light propagation in small domains.

This work was published in the journal of Biomedical Optics Express [66]. The remainder of this chapter is organized as follows. The algorithm for generating BSGs is presented in Section C.1. The necessary discretization of the FD-ERT is presented in Section C.2. Numerical phantoms presented in Section C.3 are used to validate and benchmark the algorithm. The chapter concludes with a summary of results in Section C.4 and a discussion in Section C.5.

C.1 Grid generation

The overall structure of the algorithms is shown in Fig. C.1. First the boundary information has to be provided by giving the coordinates and normal vectors of the surface that encloses the medium to be considered. Once this is given, the code determines the computational domain by filling the volume defined in the first step with grid points or mesh elements (with a user defined grid spacing Δx). Next, given certain user-defined conditions, this grid is transformed into a BSG using the block-structure mesh generator. The code then solves the light propagation problem using the discretized FD-ERT on this BSG.

To generate the appropriate BSG, we first need to determine the active computational domain given the information of the physical boundary (surface coordinates and normal vectors), which is provided as input. This process starts by defining a rectangular nominal domain that completely enclosed the arbitrarily shaped object given by the boundary information. This rectangular nominal domain is discretized on a Cartesian grid with a user specified spatial resolution $\Delta x = \Delta y = \Delta z$. This resolution has to be small enough to capture all physical effects. In our case a Δx of $1/10 \mu_s$ is usually sufficient. This method is known as the blocking-off method [79]. An overview of the major aspects of the mesh generator subroutine is presented in Fig. C.2.

The algorithm then finds all points within this nominal domain that lie inside the volume enclosed by the boundary. After discretizing Euclidean space with a fine grid, the grid points closest to the physical boundary are found (called “boundary points”) and are assigned to be part of the active region. In the next step the algorithm checks one grid point at a time and determines if it is a boundary point. When a boundary point is encountered the algorithm uses the normal vector of the surface at that grid point to determine in which direction the “inside” and “outside” regions of the medium are, relative to the current grid point. All future non-boundary points in the “inside” region of the medium are labeled as “interior points” and they are assigned to be part of the active computational domain. Points in the “outside” region of the medium are labeled “exterior points,” are inactive, and will not be used in later calculations. This labeling process is

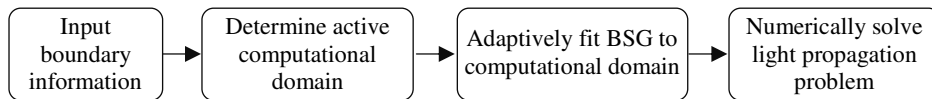


Figure C.1: General sequence in which major subroutines are executed in our algorithm.

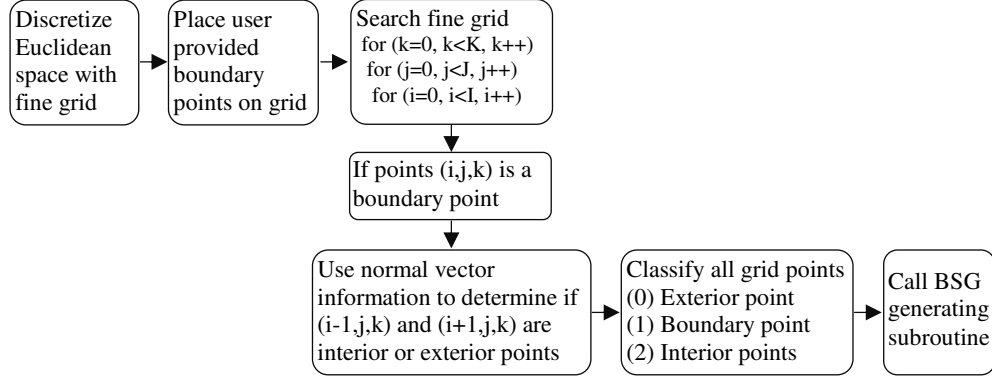


Figure C.2: Visualization of algorithm for determining computational domains.

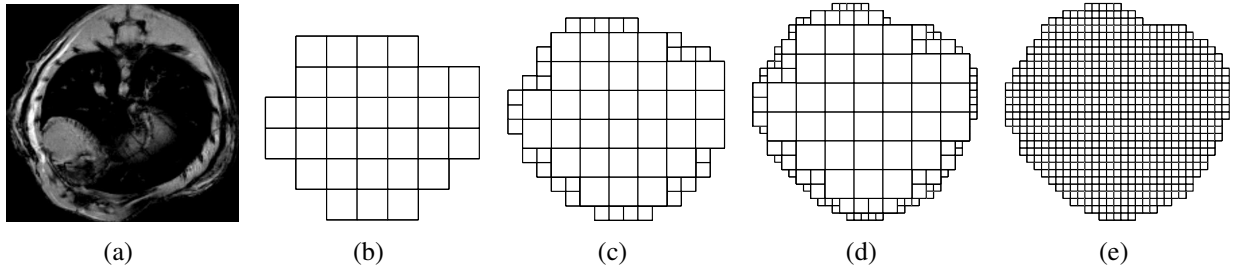


Figure C.3: Examples of the increased boundary resolution attained with BSGs relative to a single coarse grid (b,c,d) and reduction in mesh density attained by a BSG relative to a fine grid (d,e). The geometry the grids approximate is the cross-section of a mouse obtained from a segmented MRI image (a). Depicted in (b-d) are 1-, 2-, and 3-level BSGs, while (e) is the fine grid discretization of the mouse cross-section.

restarted once a new boundary point is encountered. The result of this subroutine is a Cartesian grid where all grid points are classified as exterior, boundary, or interior points.

As an example consider a cross-section through a mouse obtained from a magnetic resonance imaging (MRI) scan shown in Fig. C.3(a). A coarse discretization of the geometry defined by the tissue boundary is shown in Fig. C.3(b). The mismatch between physical and mesh boundaries is evident. It is already clear that a 2-level BSG (Fig. C.3(c)), adaptively fitted to the same cross-section, can resolve the physical boundary better than single coarse grid. A 3-level grid is able to resolve the target boundary even better (Fig. C.3(d)). Increasing the number of grid levels results in improved boundary resolution. Additionally, the number of grid points in the 3-level BSG is significantly less than the number of grid points in a single fine grid (Fig. C.3(e)).

C.2 Discretization of FD-ERT

To discretize the FD-ERT we implemented a finite-differences upwind step-method for the spatial variable and a discrete-ordinates method for the angular variable [19,67,68,69,70,71]. Here, we will first review how this is implemented on a single Cartesian grid. Subsequently, we will describe how this method is adapted to work on BSGs.

The FD-ERT is presented in detail in Section 2.3.2. For completeness, the FD-ERT and its boundary equation is reproduced below.

$$\left(\boldsymbol{\Omega} \cdot \nabla + \mu_t(\mathbf{r}) + \frac{i\omega}{v} \right) \psi(\mathbf{r}, \boldsymbol{\Omega}, \omega) = \mu_s(\mathbf{r}) \int_{4\pi} p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') \psi(\mathbf{r}, \boldsymbol{\Omega}', \omega) d\boldsymbol{\Omega}' + \frac{Q(\mathbf{r}, \omega)}{4\pi}, \quad (\text{C.1})$$

$$\psi(\mathbf{r}, \boldsymbol{\Omega}, \omega) = S(\mathbf{r}, \boldsymbol{\Omega}, \omega) + R(\boldsymbol{\Omega}' \cdot \hat{\mathbf{n}}) \psi(\mathbf{r}, \boldsymbol{\Omega}', \omega), \quad \mathbf{r} \in \partial V, \quad \boldsymbol{\Omega} \cdot \hat{\mathbf{n}} < 0. \quad (\text{C.2})$$

C.2.1 Discretization on single grid

The first step in discretizing the ERT was to use the discrete-ordinates method to replace the integral term with the extended trapezoidal rule [19]. This approximation was given by Eq. C.3, where k is the ordinate number, ψ_k is the radiance in the k th ordinate, and ω_k is a predetermined ordinate weight with full level symmetry [70, 71]. The integral term does not require special treatment for implementation on BSGs.

$$\int_{4\pi} \psi(\mathbf{r}, \boldsymbol{\Omega}) d\boldsymbol{\Omega} = \sum_{k=1}^K \omega_k \psi_k(\mathbf{r}) \quad (\text{C.3})$$

The partial derivative terms were discretized using the upwind-step method. The directional cosines of a given ordinate determined the upwind-direction for that particular ordinate. There were eight possible numerical schemes, one for each octant in the three dimensional Cartesian coordinate system. For example, when all directional cosines are positive the upwind-step method requires an Euler step in the negative x -, y -, and z -axis. For this example, the discretization of (C.1) and (C.2) is given by (C.4) and (C.5), respectively.

$$\begin{aligned}
& \Omega_x \frac{[\psi]_{ijlk} - [\psi]_{(i-1)jlk}}{\Delta x} + \Omega_y \frac{[\psi]_{ijlk} - [\psi]_{i(j-1)lk}}{\Delta y} + \Omega_z \frac{[\psi]_{ijlk} - [\psi]_{ij(l-1)k}}{\Delta z} \\
& + [\mu_t]_{ijl} [\psi]_{ijlk} + \frac{\omega_m i}{\nu} [\psi]_{ijlk} = [\mu_s]_{ijl} \sum_{k'=1}^K \omega_{k'} p_{kk'} [\psi]_{ijlk'} + \frac{Q_{ijl}}{4\pi}
\end{aligned} \tag{C.4}$$

$$[\psi]_{ijlk} = S_k + R(\Omega_{k'} \cdot n_{ijl}) [\psi]_{ijlk}, \quad \Omega_k \cdot n_{ijl} < 0 \tag{C.5}$$

The term $p_{kk'}$ in (C.4) is the *Henyey-Greenstein* phase function and is given by (C.6), where g is the anisotropy factor.

$$p_{kk'} = \frac{1 - g^2}{4\pi (1 + g^2 - 2g\Omega_k \cdot \Omega_{k'})^{3/2}} \tag{C.6}$$

The radiance, ψ , can be solved from (C.4) to (C.5) with any number of established algorithms. In this work we implement the source iteration technique (i.e. the matrix-free point-wise Gauss-Seidel Method).

The fluence is given by

$$\hat{\phi}(\mathbf{r}, \omega_m) = \int_{4\pi} \hat{\psi}(\mathbf{r}, \boldsymbol{\Omega}, \omega_m) d\boldsymbol{\Omega}, \tag{C.7}$$

and is discretized through application of the extended trapezoidal rule [67, 70, 71, 206, 218], and is given by

$$\phi_{ijl} = \sum_{k=1}^K \omega_k \psi_{ijlk} \tag{C.8}$$

C.2.2 Discretization of FD-ERT on BSGs

To solve the FD-ERT on BSGs, the Euler step used in the step-method (Eq. C.4) had to be changed to a step of variable size and became dependent on the local grid. For example, Δx became Δx_{ijl} where ijl denoted the current grid point and its value was determined by the size of the local grid. Implementing a

finite differences numerical scheme with a variable Euler step in a matrix-free formulation on BSGs was a complicated endeavor and requires great care. The main difficulty arises when solving for the FD-ERT on mesh points on an interior grid boundary (i.e. points that straddled both a coarse grid and a fine grid region).

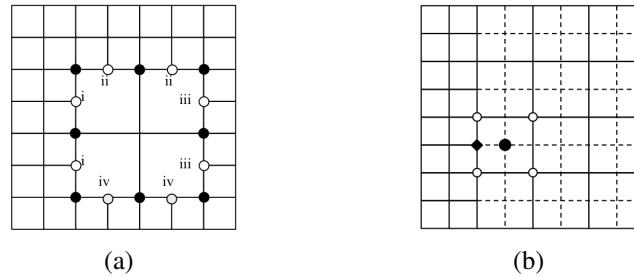


Figure C.4: (a) Example of possible grid points on interior boundaries. There are five distinct cases. The first four cases are labeled i-iv and the fifth case is specified by the black dot. (b) The grid point denoted by the black diamond is an interior boundary point; the fine grid is to its left and the coarse grid is to its right. The black dot is the virtual point that must be created when the differencing scheme in the x-direction requires a forward Euler-step. Solutions at the four white dots surrounding the black dot are averaged to create the virtual point.

Mesh points on a boundary between fine and coarse grids required special attention because the points necessary to complete the numerical stencil were not always present. The numerical stencil requires four neighbors in two-dimensions and six neighbors in three-dimensions. Mesh points on an interior grid boundary do not always have a full set of neighboring mesh points to complete the numerical stencil. This problem can be overcome by adjusting the numerical scheme for each individual boundary point, or by creating the missing point so that the normal scheme is applicable. In this work we consider the latter option. The missing point (virtual point) is constructed through interpolation using neighboring points.

For illustration consider the 2-dimensional example in Fig. C.4. Here five cases must be considered independently for a given octant. For example, when all directional cosines are positive, points ii and iii will require creating a virtual point interior to the coarse grid. However, points i and iv have all neighbors necessary to complete the stencil (Fig. C.4(a)). The fifth case (black dots in Fig. C.4(a)) can be treated as points on the fine grid or on the coarse grid. By treating these boundary points as coarse grid points we assure that they will always have a complete set of neighbors and no further special treatment is necessary. The set of points used to create the virtual point varies according to the type of boundary point. The different types of boundary points can be reduced to five in two dimensions (Fig. C.4(a)) and nineteen in three dimensions.

There are many ways to interpolate values for the virtual points needed to complete the differencing

schemes. Possible interpolation methods include averaging of the nearest four neighbors, as well as bilinear, biquadratic, and bicubic interpolation [206]. Higher order interpolation schemes are typically preferred because they are able to better resolve the curvature of the solution field. However, higher order methods are computationally more expensive. Therefore, the type of interpolation scheme to be used must be chosen based on accuracy and computational cost.

In this work we implement the four-neighbor averaging scheme because it yields sufficiently accurate results. The solutions from the missing point's four nearest neighbors (Fig. C.4(b)) are averaged to create the virtual point.

C.3 Numerical phantoms

The accuracy and computation speed of the BSG algorithm was compared to a previously published single grid algorithm that also uses a combination of discrete ordinates and the finite differences upwind-step method to solve the FD-ERT [19]. The performance of the BSG algorithm was tested on multiple numerical phantoms. For brevity, only results from two phantoms are presented: (1) a 2 cm diameter disk and (2) a homogenous three-dimensional mouse phantom obtained from an MRI data set (Fig. C.5). We solved the FD-ERT on these phantoms on single Cartesian grids, on both coarse and fine grids, and on BSGs. We evaluated the performance of the BSG algorithm by comparing these solutions. The specific optical properties used are summarized in Tables C.1 and C.2.

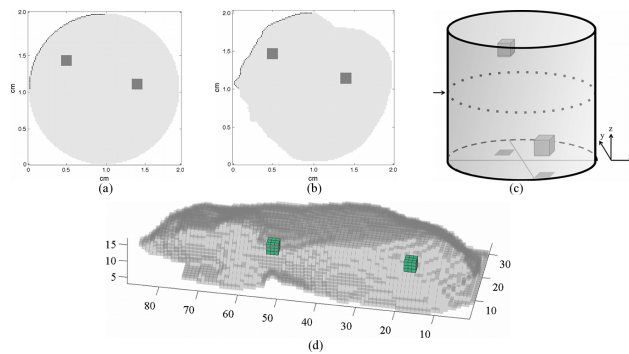


Figure C.5: (a) disk, (b) 2D mouse cross section, (c) cylinder, and (d) 3D mouse model. The positions of boundary sources are shown for as black dots, while detectors are assigned to every boundary point. (d) Sources for the mouse are located within the interval defined by $x = [16, 18]$, $y = [43, 46]$, and $z = 0$. Similarly, detectors are defined within $x = [16, 18]$, $y = [43, 46]$, and $z = 15$.

Table C.1: Properties of embedded fluorescent probes.

Property	Value
Quantum Yield	0.93%
Life Time	5.0 ns
Absorption	0.4 cm^{-1}
Dimensions	$0.125 \times 0.125 \text{ cm}^{-1}$

Table C.2: Summary of parameter values used for simulations.

Case	ω [MHz]	μ_a [cm^{-1}]	μ_s [cm^{-1}]	n
1	100.00	0.1	10.0	1.00
2	100.00	0.2	10.0	1.00
3	100.00	0.1	5.0	1.00
4	100.00	0.1	10.0	1.37
5	200.00	0.1	10.0	1.00

The disk phantom is of interest because it represents an instance where a coarse grid cannot place a single grid point on the true phantom surface, whereas a finer grid can always better approximate the true boundary. This is a worst-case scenario for the coarse grid. The anatomically correct phantom is of interest because it is representative of the arbitrarily shaped geometries encountered in practice where boundaries can have convex and concave regions. Generating a mesh for the anatomically correct phantom and solving the light propagation problem is a good test for both the mesh-generating routine and the FD-ERT numerical solver.

As can be seen from Fig. C.5(a), the disk phantom has two embedded fluorescent probes. The properties of the fluorescent probes are listed in Table C.1. The frequency modulated boundary sources are defined on the top-left quadrant of the disk (represented by black dots). Thus, instead of defining a single boundary source we define a constant source area. This is important because the number of boundary points on a given model increases with decreasing Δx . With this setup we ensure the number of photons injected into the phantom is independent of the number of boundary points. The source density is $8 \times 10^9 \text{ photons cm}^{-2} \text{ sr}^{-1}$. The optical properties of the disk phantom are varied and they are summarized in Tables C.1 and C.2.

Figure C.5(b) shows the two-dimensional mouse-like phantom. As in the disk-shaped phantom, this phantom has two embedded fluorescent probes. The boundary sources for excitation are on the top-left quadrant and are defined as external boundary sources for reasons similar to those presented for the disk

phantom. Source density is 8×10^9 photons $\text{cm}^{-2} \text{sr}^{-1}$.

The grid spacing used for simulations in two dimensions was $\Delta x = 2/256, 2/128, 2/64, 2/32$ cm (where $\Delta x = \Delta y$). For each grid refinement, we determined the solution to the ERT on 1-, 2-, and 3-level BSGs and compared them to the benchmark solution. The benchmark solution for these simulations was the solution on the finest single grid ($\Delta x = 2/256$ cm). In addition, the optical parameters (μ_a, μ_s), modulation frequency (ω), and refractive index (n_m) of each phantom were varied.

Simulations on three-dimensional phantoms (Fig. C.5(c,d)) were carried out using parameters from case 4 in Table C.2. The benchmark solutions for the cylinder and mouse phantoms are defined on grids with $\Delta x = 2/128$ cm and $\Delta x = 0.1$ cm, respectively ($\Delta x = \Delta y = \Delta z$).

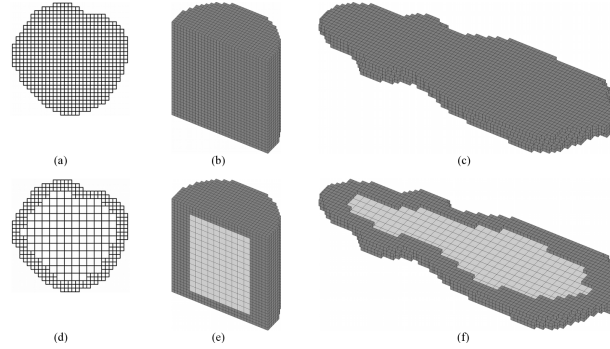


Figure C.6: (a-c) Examples of single structured grids fitted to the disk, cylinder, and mouse phantom, respectively. (d-f) 2-level BSGs fitted to same three phantoms. Three-dimensional BSGs and their interior structure are visualized by showing only a section of the full three-dimensional shape (e,f).

C.4 Results

The results from this study are briefly summarized. Examples of BSGs fitted to two- and three-dimensional phantoms are shown to demonstrate that the BSG generator worked properly (Fig. C.6). The computation time required for each forward problem and the accuracy of the solution at the boundary using the mean percent error (MPE) measure was computed. The results presented in this section focus on case 4 (Table C.2) only. The cited grid spacing (Δx) always refers to the spacing of the finest grid (i.e. the grid near the boundaries). As an example, a 2-level BSG with grid spacing Δx has an embedded section of coarse grid points whose spacing is $2\Delta x$. The terms 1L, 2L, and 3L refer to the number of grid levels in the BSGs. Computation time is

quantified as relative speed up (RSU) compared to the benchmark.

Table C.3: Computation time, MPE, and RSU on a mouse cross-section.

Case	Δx [cm]	Time [s]			MPE [%]			RSU		
		1L	2L	3L	1L	2L	3L	1L	2L	3L
1	2/32	10.7	4.9	–	25.3	60.3	–	–	1.2	–
	2/64	58.5	19.4	13.8	3.6	7.1	20.3	–	2.0	3.2
	2/128	305.3	98.1	59.1	1.2	2.1	4.5	–	2.1	4.2
	2/256	1453.8	523.0	380.3	0.0	0.2	0.8	–	1.8	2.8
2	2/32	8.9	4.3	–	108.9	444.4	–	–	1.1	–
	2/64	47.9	16.3	11.8	6.9	24.6	127.7	–	1.9	3.0
	2/128	235.5	77.8	51.1	2.0	5.6	18.3	–	2.0	3.6
	2/256	1134.3	405.6	294.2	0.0	0.9	2.3	–	1.8	2.9
3	2/32	4.7	2.3	–	4.4	6.5	–	–	1.0	–
	2/64	24.4	8.5	6.5	1.2	1.6	2.9	–	1.9	2.7
	2/128	119.3	40.4	26.8	0.5	0.6	0.8	–	2.0	3.5
	2/256	559.1	201.8	149.1	0.0	0.3	0.2	–	1.8	2.8
4	2/32	11.8	5.6	–	43.9	102.1	–	–	1.1	–
	2/64	62.4	21.2	15.6	6.2	14.7	46.8	–	2.0	3.0
	2/128	308.9	101.0	65.5	1.5	3.2	7.1	–	2.1	3.7
	2/256	1558.3	566.9	405.4	0.0	0.3	1.2	–	1.7	2.8
5	2/32	10.1	4.8	–	27.3	62.7	–	–	1.1	–
	2/64	53.8	18.5	13.3	3.6	7.4	21.1	–	1.9	3.0
	2/128	266.5	88.4	57.5	1.2	2.1	4.6	–	2.0	3.6
	2/256	1275.9	458.9	331.8	0.0	0.2	0.8	–	1.8	2.8

The results for the mouse cross-section phantom are summarized in Table C.3. Solutions to the excitation problem using a 2-level BSG were obtained about 2 times faster than the benchmark with a relatively small increase in error. The notable exceptions to these observations are the results from 2-level BSGs with $\Delta x = 2/32$ cm. In these cases the solution on 2-level BSGs is obtained only 1 time faster than the single grid solution and the increase in error is very large. This may occur because the interior of the BSG is very coarse ($\Delta x = 2/16$ cm), which may lead to large numerical error.

Comparing case 1 and case 4 was particularly interesting because the results are significantly different while the only difference in the simulations was the use of refractive index mismatch in case 4. In case 1, when there was no refractive index mismatch at the phantom surface, the error in the solution computed on the coarsest grid was 25.3%. However, in case 4, this same error increased to 43.9% when the phantom was assigned a refractive index of $n_m = 1.37$. This is evidence that a finer mesh is required when taking into

account the refractive index mismatch at the tissue boundary.

As expected, the performance of the algorithm with the given set of grid discretization is heavily dependent on scattering and absorption parameters. For example, the error in the solution computed on coarse grids is particularly large when the absorption coefficient is large (case 2). However, the error is relatively small even on the coarse grid when the scattering coefficient is very small (case 3). We note that the performance of the algorithm is not dependent on modulation frequency (case 5).

Representative examples of solutions to the FD-ERT on the cylindrical and three-dimensional mouse phantoms are shown in Fig. C.7. Results from simulations on the cylindrical phantom are summarized in Table C.4. The error in the solution computed on a 2-level grid ($\Delta x = 2/128$ cm) is only 0.28%, while the solution computed on a single coarse grid ($\Delta x = 2/64$ cm) 2.94%. Thus, solving the FD-ERT on a 2-level grid instead of a coarse grid reduced the error by 2.66%. Similarly, the solution computed on a 3-level grid ($\Delta x = 2/128$ cm) is 30.21%, while the error in the solution computed on the coarsest grid ($\Delta x = 2/32$ cm) is 78.25%. The error in the solution is reduced by 48.29%. In addition, using BSGs reduces computation time. Solutions on 2- and 3- level grids are obtained 1.5 and 3.0 times faster than the solution on the fine grid, respectively. Results from simulations on the three-dimensional mouse phantom are similar.

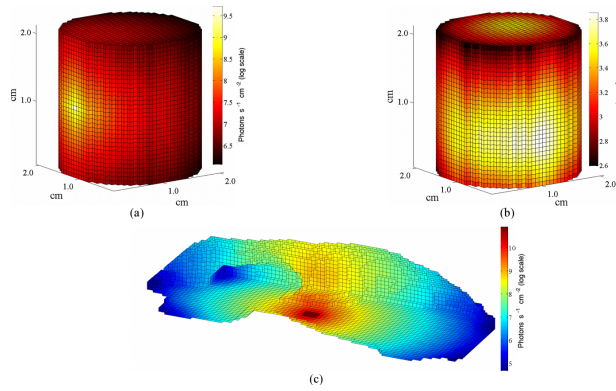


Figure C.7: Three-dimensional representations of excitation (a) and emission (b) fluence on the cylindrical phantom. (c) Sample excitation on the three-dimension mouse phantom.

C.5 Discussion

The motivation for solving the FD-ERT on BSGs came from the need to reduce computation time without sacrificing accuracy of numerical solutions to the FD-ERT on structured grids. The error in the solutions to

Table C.4: Computation time, MPE, and RSU on a cylindrical phantom

Case	Δx [cm]	Time [s]			MPE [%]			RSU		
		1L	2L	3L	1L	2L	3L	1L	2L	3L
4	2/32	394.5	201.2	–	78.25	104.31	–	–	0.96	–
	2/64	5464.15	2068.4	1054.2	2.94	5.06	80.25	–	1.64	4.18
	2/128	68489.96	27892.7	17233.3	0	0.28	30.21	–	1.46	3.0

the light propagation problem obtained on structured grids come from (1) inherent truncation errors from approximating a continuous equation with a discrete numerical scheme and (2) a mismatch between the numerical and physical boundary. Finely discretized grids can accurately model complex boundaries but solving the ERT on these grids is computationally expensive. Solving the ERT on coarse grids requires less computational effort; however, the accuracy of the solution can be very poor. A compromise between fine grids and coarse grids was the use of BSGs as presented in this paper. The interior of a BSG is primarily a coarse grid, while the outer layers of the grid, which are closer to the physical boundary, are more finely discretized.

In this work we presented the first algorithm for solving the FD-ERT on BSGs. This is also the first algorithm that incorporates reflective boundary conditions and a subroutine for generating BSGs directly into the algorithm for solving the FD-ERT. The BSG generator uses boundary information to determine the computational domain, discretizes it with a fine grid, then it adaptively coarsens the grid up to a user defined level. The final computation grid is a union of fine and coarse grids, where the coarse grid is restricted to the interior and the grid spacing becomes smaller near the boundary. The ERT is solved on this grid with a combination of the upwind step method and the discrete ordinates approximation. We use the blocked-off region method to treat curved boundaries.

Solutions to the ERT computed on single structured grids are corrupted by inherent error of the numerical approximation to the continuous equation and error due to poorly resolved boundaries. The inherent numerical error arises from the first order upwind step method approximation to the derivative terms and is proportional to spatial discretization ($\Delta x =$). In addition, there is numerical error due to the S_N approximation to the integral terms (it decreases with increasing order of the S_N method). Error due to poorly resolved boundaries arises when the single Cartesian grid does not accurately approximate the physical boundary. This error is particularly large when a coarse grid is used to approximate curved geometries.

The total error associated with solutions to the ERT on BSGs is similar to the errors associated with solutions on a single coarse grid; however, there are two fundamental differences: (1) boundary errors are reduced because refining the grid near the boundary more accurately captures boundary effects and (2) a new source of error is introduced by interpolating at boundaries between coarse and fine grid sections inside the computational domain. The inherent numerical error associated with using the upwind step method on BSGs is similar to the numerical error expected from using this scheme on a single coarse grid because the majority of the BSG is composed of coarse grid points.

The need for BSGs becomes apparent when geometries are arbitrarily shaped. Results from 2D and 3D mouse phantoms confirm that the error due to a poorly resolved boundary is significant, especially when the refractive index mismatch at the air/tissue interface is taken into account. In this case, we found that the error in the solution computed on a single coarse grid was 25.3% when the refractive index mismatch was not taken into account. However, the error increased to 43.9% when there was refractive index mismatch at the boundary.

Simulations on a disk and mouse phantom show that solutions on BSGs are always obtained faster than the corresponding fine grid solutions. Solutions on 2- and 3-level BSGs were obtained in about 1/3 and 1/4 the time it took to obtain the same solution on a single fine grid, respectively. In general, the speed up achieved by the algorithm was not affected by changes in optical properties, refractive index, or modulation frequency.

Through analysis of simulations on the phantoms we show that solutions on BSGs are significantly more accurate than solutions on single coarse grids. Increasing the refinement of the grid near the boundary decreased the overall error in the solution for all cases studied. The general trend can be summarized as follows: the MPE of partial current measurements from solutions computed on 2- and 3-level BSGs are reduced to about 1/3 and 1/6, respectively, from the MPE of a solution obtained on a single coarse grid.

Overall we find that solving the FD-ERT on BSGs yields an attractive algorithm for modeling the light propagation problem on geometries with arbitrary shape without using dense and finely discretized Cartesian grids. The algorithm provides a method to substantially reduce the error due to poorly resolved boundaries. Furthermore, solutions to the FD-ERT on BSG are obtained at a lower computation cost compared to solutions computed on a single fine grid.

Overall, this method of solving the ERT is still burdened with high computational demand, motivating

the need for using a more economical light propagation model, such as the SP_N approximation to the ERT.