NONLINEAR MULTIGRID INVERSION ALGORITHMS WITH APPLICATIONS TO STATISTICAL IMAGE RECONSTRUCTION

A Thesis

Submitted to the Faculty

of

Purdue University

by

Seungseok Oh

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2005

To my parents and Suna

ACKNOWLEDGMENTS

I was fortunate enough to have not just one but two exceptional advisors, Professor Charles Bouman and Professor Kevin Webb. I thank them for their guidance, mentoring, career advice, and endless patience. Studying under the guidance of two advisors was demanding, but at the same time rewarding: each of them has his own perspective, his own expertise field, his own philosophy, and his own style. They made me realize the importance of collaborative, interdisciplinary research as well as uncompromising academic standards.

I am grateful to the other committee members, Professor Peter Doerschuk and Professor Bradley Lucier, for their helpful suggestions. I am also grateful to Professor Rick Millane for fruitful discussion and thorough reading of a chapter. I thank Professor Jan Allebach for his invaluable advice, which helped me advance my career objectives. I would also like to express gratitude to Adam Milstein. I enjoyed our fruitful collaboration that resulted in co-authorship of our work.

I wish to express gratitude to my parents who, throughout my life, have always offered unconditional love and support to me.

Finally, I would like to thank my wife, Suna, for her endless love, encouragement, support, patience, and her endearing smile.

TABLE OF CONTENTS

Page

LIST OF TABLES v

LIST OF FIGURES vi

ABSTRACT x

1 INTRODUCTION 1

2 A GENERAL FRAMEWORK FOR NONLINEAR MULTIGRID INVERSION 3

2.1 Introduction 3

2.2 Multigrid Inversion Framework 6

2.2.1 Inverse problems 7

2.2.2 Fixed-grid inversion 8

2.2.3 Multigrid inversion algorithm 9

2.2.4 Convergence of multigrid inversion 16

2.2.5 Stabilizing functionals 18

2.3 Application to Optical Diffusion Tomography 19

2.4 Numerical Results 24

2.4.1 Evaluation of required forward model resolution 25

2.4.2 Multigrid performance evaluation 29

2.5 Conclusions 32

3 MULTIGRID TOMOGRAPHIC INVERSION WITH VARIABLE RESOLUTION DATA AND IMAGE SPACES 37

3.1 Introduction 37

3.2 Multigrid Inversion with Variable Resolution Data and Image Spaces 40

3.2.1 Quadratic data term case 40

3.2.2 Poisson data case 44

Page

3.3 Adaptive Computation Allocation 48
3.4 Applications to Bayesian Emission and Transmission Tomography . . 50
3.4.1 Multigrid tomographic inversion with quadratic data term . . 50
3.4.2 Multigrid tomographic inversion for Poisson data model 52
3.5 Numerical Results 53
3.6 Conclusions 57
4 SOURCE-DETECTOR CALIBRATION IN THREE-DIMENSIONAL BAYESIAN OPTICAL DIFFUSION TOMOGRAPHY 66
4.1 Introduction 66
4.2 Problem Formulation 68
4.3 Optimization 72
4.4 Results 77
4.4.1 Simulation 77
4.4.2 Experiment 90
4.5 Conclusions 94
LIST OF REFERENCES 95
A PROOF OF MULTIGRID MONOTONE CONVERGENCE 104
B COMPUTATIONAL COMPLEXITY OF MULTIGRID INVERSION . . . 107
C COMPUTATIONAL COMPLEXITY OF MULTIGRID INVERSION WITH VARIABLE DATA RESOLUTION 110
D MULTIGRID INVERSION WITH VARIABLE DATA RESOLUTION FOR GAUSSIAN DATA WITH NOISE SCALING PARAMETER ESTIMATION 113
VITA 117

LIST OF TABLES

Table Page

2.1 Distortion-to-noise (DNR) ratio for various forward model resolutions. Coarse discretization increased forward model error, and source/detector pairs on the same face had much higher DNR. 26

2.2 The normalization parameter σ that yields the best reconstruction and the resulting RMS image error between the reconstructions and the decimation of the true phantom. 28

2.3 Complexity comparison for each algorithm. Theoretical complex multiplications are estimated with (B.1) and theoretical relative complexity is the ratio of the required number of multiplications for one iteration to that for one fixed-grid iteration. Experimental relative complexity is the ratio of user time required for one iteration to that for one fixed-grid iteration. 31

LIST OF FIGURES

Figure Page

2.1 The role of adjustment term $r^{(q+1)}x^{(q+1)}$. (a) When the gradients of the fine scale and coarse scale cost functionals are different at the initial value, the updated value may increase the fine grid cost functional's value. (b) When the gradients of the two functionals are matched, a properly chosen coarse scale functional can guarantee that the coarse scale update reduces the fine scale cost. 12

2.2 Pseudo-code specification of a two-grid inversion algorithm. The notation $c^{(q+1)}(x^{(q+1)}; y^{(q+1)}, r^{(q+1)})$ is used to make the cost functional's dependency on $y^{(q+1)}$ and $r^{(q+1)}$ explicit. 14

2.3 Pseudo-code specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion. The Multigrid-V algorithm is similar to the 2-grid algorithm, but recursively calls itself to perform the coarse grid update. 15

2.4 Pseudo-code specification of fixed grid and multigrid inversion methods for the ODT problem showing (a) main routine for ODT problems, (b) fixed-grid update, and (c) Multigrid-V inversion. 23

2.5 (a) Source and (b) detector pattern on each face of the cube geometry. Two data set scenarios were considered: one containing all source/detector pairs, and a second containing only source/detector pairs on different faces. 26

2.6 A cross-section through (a) the inhomogeneous phantom, and the best reconstructions obtained using source detector pairs on different faces with (b) $65\times65\times65$ resolution, (c) $33\times33\times33$ resolution, (d) $17\times17\times17$ resolution, and (e) all source detector pairs with $65 \times 65 \times 65$ resolution. 27

2.7 Convergence of (a) cost function and (b) RMS image error when reconstructions were initialized with average values of true phantom. All multigrid algorithms converge about 13 times faster than the fixed-grid algorithm. 33

Figure Page

2.8 Cross-sections of reconstructions on the plane through the centers of the inhomogeneities using (a) 4 level multigrid with 19.35 iterations, (b) 3 level multigrid with 19.95 iterations, (c) 2 level multigrid with 18.24 iterations, and (d) 270 fixed grid iterations. All the multigrid reconstructions have better image quality the the fixed grid reconstruction. 34

2.9 Convergence of (a) cost function and (b) RMS image error with a poor initial guess. For higher level multigrid algorithms, the convergence was faster. In particular, the four level multigrid algorithm converged almost as fast as when the reconstruction was initialized with the true phantom's average value. 35

3.1 Pseudo-code specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion. 45

3.2 Adaptive multigrid-V scheme 49

3.3 (a) true phantom (b) CBP reconstruction for emission tomography (c) CBP reconstruction for transmission tomography 55

3.4 Convergence in emission tomography with quadratic data term in terms of (a) cost function and (b) image rms error 58

3.5 Convergence in emission tomography with the Poisson noise model in terms of (a) cost function and (b) image rms error 59

3.6 Convergence in transmission tomography with quadratic data term in terms of (a) cost function and (b) image rms error 60

3.7 Convergence in transmission tomography with the Poisson noise model in terms of (a) cost function and (b) image rms error 61

3.8 Reconstructions for emission tomography with quadratic data term: fixed-grid algorithm with (a) 7 iterations (b) 14 iterations (c) 28 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (7.79 iterations); and (f) multigrid algorithm with variable data resolution (5.94 iterations) 62

3.9 Reconstructions for emission tomography with the Poisson noise model: fixed-grid algorithm with (a) 7 iterations (b) 14 iterations (c) 28 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (8.06 iterations); and (f) multigrid algorithm with variable data resolution (5.31 iterations) 63

Figure Page

3.10 Reconstructions for transmission tomography with quadratic data term: fixed-grid algorithm with (a) 7 iterations (b) 14 iterations (c) 28 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (7.48 iterations); and (f) multigrid algorithm with variable data resolution (5.81 iterations) 64

3.11 Reconstructions for transmission tomography with the Poisson noise model: fixed-grid algorithm with (a) 8 iterations (b) 16 iterations (c) 32 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (9.06 iterations); and (f) multigrid algorithm with variable data resolution (6.46 iterations) 65

4.1 Pseudo-code specification for (a) the overall optimization procedure and (b) the image update by one ICD scan. 76

4.2 Isosurface plots (at 0.04 cm^{-1} for μ_a, and 0.02 cm for D) for μ_a (left column) and D (right column) for Phantom A: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration. 79

4.3 Cross-sections through the centers of the inhomogeneities (z=0.5 cm for μ_a, z=1.5 cm for D) for μ_a (left column) and D (right column) of Phantom A: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration. 80

4.4 Isosurface plots (at 0.04 cm^{-1} for μ_a, and 0.02 cm for D) for μ_a (left column) and D (right column) for Phantom B: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration. 81

4.5 Cross-sections through the centers of the inhomogeneities (z=0.0 cm for μ_a, z=0.25 cm for D) for μ_a (left column) and D (right column) of Phantom B: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration. 82

4.6 (a) Locations of sources and detectors, (b) Several levels of boundaries: zero-flux boundary, physical boundary, source-detector boundary, and imaging boundary, from the outer boundary. 83

4.7 (a) Source/detector coupling coefficients used in the simulations. The estimation error of coupling coefficients for (b) Phantom A and (c) Phantom B after 30 iterations. Note that the scale of (b) and (c) is 10 times of that of (a). 84

Figure Page

4.8 The normalized root mean square error between the phantom and the reconstructed images for (a) Phantom A and (b) Phantom B. 85

4.9 (a) RMS error in the estimated coupling coefficients versus iteration. (b) Convergence of coupling coefficients for Group 1 (—) and Group 2 (- - -) for Phantom B. 86

4.10 Image NRMSE comparison between the reconstruction with coupling coefficient calibration and the reconstruction with coupling coefficients fixed to $1 + 0i$, for various standard deviations of coupling coefficients. Images were obtained after 30 iterations. 88

4.11 Cross-sections of the reconstructed images through the centers of the inhomogeneities (z=0.5 cm for μ_a, z=1.5 cm for D) : for $\sigma_{\text{coeff}} = 0.02$ for (a) μ_a and (b) D, and for $\sigma_{\text{coeff}} = 0.04$ for (c) μ_a and (d) D. 89

4.12 (a) Culture flask with the absorbing cylinder embedded in a scattering Intralipid solution. (b) Schematic diagram of the apparatus used to collect data. 92

4.13 Cross-sections for reconstructed images of an absorbing cylinder with (a) two complex valued calibration coefficients, (b) a single complex calibration coefficient, (c) a single real calibration coefficient, and (d) all calibration coefficients assumed to be 1. 93

C.1 Comparison between the theoretical complexity and the measure CPU time for the multigrid algorithms with (a) fixed data resolution and (b) variable data resolution 112

D.1 Pseudo-code specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion for Gaussian data with unknown noise scaling parameter estimation 116

ABSTRACT

Oh, Seungseok. Ph.D., Purdue University, May, 2005. Nonlinear multigrid inversion algorithms with applications to statistical image reconstruction. Major Professors: Charles A. Bouman and Kevin J. Webb.

Many tasks in image processing applications, such as reconstruction, deblurring, and registration, depend on the solution to inverse problems. In this thesis, we present nonlinear multigrid inversion methods for solving computationally expensive inverse problems. The multigrid inversion algorithm results from the application of recursive multigrid techniques to the solution of optimization problems arising from inverse problems. The method works by dynamically adjusting the cost functionals at different scales so that they are consistent with, and ultimately reduce, the finest scale cost functional. In this way, the multigrid inversion algorithm efficiently computes the solution to the desired fine scale inversion problem.

While multigrid inversion is a general framework applicable to a wide variety of inverse problems, it is particulary well-suited for the inversion of nonlinear forward problems such as those modeled by the solution to partial differential equations since the new algorithm can greatly reduce computation by more coarsely descretizing both the forward and inverse problems at lower resolutions. An application of our method to optical diffusion tomography shows the potential for very large computational savings, better reconstruction quality, and robust convergence with a range of initialization conditions for this non-convex optimization problem.

The method is extended to further reduce computations by reducing the resolutions of the data space as well as the parameter space at coarse scales. Applications of the approach to Bayesian reconstruction algorithms in transmission and emission tomography are presented, both with a Poisson noise model and with a quadratic

data term. Simulation results indicate that the proposed multigrid approach results in significant improvement in convergence speed compared to the fixed-grid iterative coordinate descent (ICD) method and a multigrid method with fixed data resolution.

1. INTRODUCTION

Many tasks in image processing applications, such as reconstruction, restoration, registration, and analysis, may be formulated as inverse problems. Often, the numerical solution of these inverse problems can be computationally demanding. In this thesis, we propose a general framework for nonlinear multigrid inversion that is applicable to a wide variety of inverse problems, and we describe its applications to Bayesian image reconstruction for diffusion tomography, transmission tomography, and emission tomography.

Chapter two presents a general framework for nonlinear multigrid inversion and discusses its convergence. Our multigrid inversion framework results from the application of recursive multigrid techniques to the solution of optimization problems arising from inverse problems. The method works by dynamically adjusting the cost functionals at different scales so that they are consistent with, and ultimately reduce, the finest scale cost functional. A sufficient condition for monotone convergence of the multigrid optimization is proved. We apply the multigrid approach to optical diffusion tomography (ODT), which requires the inversion of a forward problem that is modeled by the solution to a partial differential equation. An application of our method to Bayesian ODT with a generalized Gaussian Markov random field (GGMRF) image prior model demonstrates the potential for very large computational savings, better reconstruction quality, and robust convergence with a range of initialization conditions.

Chapter three extends the multigrid approach to change the dimensions of the data space as well as the parameter space, thus further reducing computation. Its advantage is particularly important for conventional tomography, such as X-ray computed tomography (CT) and positron emission tomography (PET), where observa-

tion resolutions may differ for different scales. In addition, to further improve computational efficiency, computations are adaptively allocated to the scale at which the algorithm can best reduce the cost. Its applications to Bayesian reconstruction algorithms for CT and PET with a GGMRF image prior are presented both for an exact Poisson measurement noise model and for an approximate Gaussian one.

The last topic of this thesis is a statistical estimation approach for calibrating ODT data collection systems. Unknown optical source and detector coupling is modeled with complex-valued coupling coefficients embedded in a data likelihood function in a Bayesian framework, and the coefficients and image are simultaneously estimated. Simulation and experimental results show that our method can substantially improve reconstruction quality with no prior reference measurement.

2. A GENERAL FRAMEWORK FOR NONLINEAR MULTIGRID INVERSION

2.1 Introduction

A large class of image processing problems, such as deblurring, high-resolution rendering, image recovery, image segmentation, motion analysis, and tomography, require the solution of inverse problems. Often, the numerical solution of these inverse problems can be computationally demanding, particularly when the problem must be formulated in three dimensions.

Recently, some new imaging modalities, such as optical diffusion tomography (ODT) [1–4] and electrical impedance tomography (EIT) [5], have received much attention. For example, ODT holds great potential as a safe, non-invasive medical diagnostic modality with chemical specificity [6]. However, the inverse problems associated with these new modalities present a number of difficult challenges. First, the forward models are described by the solution of a partial differential equation (PDE) which is computationally demanding to solve. Second, the unknown image is formed by the coefficients of the PDE, so the forward model is highly nonlinear, even when the PDE is itself linear. Finally, these problems typically are inherently 3-D due to the 3-D propagation of energy in the scattering media being modeled. Since many phenomena in nature are mathematically described by PDEs, numerous other inverse problems have similar computational difficulties, including microwave tomography [7], thermal wave tomography [8], and inverse scattering [9].

To solve inverse problems, most algorithms, such as conjugate gradient (CG), steepest descent (SD), and iterative coordinate descent (ICD) [10] work by performing all computations using a fixed discretization grid. While tremendous progress has

been made in reducing the computational complexity of these fixed grid methods, computational cost is still of great concern. Perhaps more importantly, fixed grid optimization methods are essentially performing a local search of the cost function, and are therefore more susceptible to being trapped in local minima that can result in poorer quality reconstructions.

Multiresolution techniques have been widely investigated to reduce computation for inverse problems. Even simple multiresolution approaches, such as initializing fine resolution iterations with coarse solutions [11–15], have been shown to be effective in many imaging problems. Wavelets have been studied for Bayesian tomography [16–20], and both wavelet and multiresolution models have been applied in Bayesian formulations of emission tomography [21–24] and thermal wave tomography [25]. For ODT, a two resolution wavelet decomposition was used to speed inversion of a problem linearized with a Born approximation [26].

Multigrid methods are a special class of multiresolution algorithms which work by recursively operating on the data at different resolutions, using the ideas of nested iterations and coarse grid correction [27–32]. Multigrid algorithms originally attracted interest as a method for solving PDEs by effectively removing smooth error components, which are not always damped in fixed-grid relaxation schemes. In particular, the full approximation scheme (FAS) of Brandt [27] can be used to solve nonlinear PDEs. Multigrid methods have been used to expedite convergence in various image processing problems, for example, lightness computation [33], shape-from-X [33, 34], optical flow estimation [33, 35–38], signal/image smoothing [39, 40], image segmentation [40, 41], image matching [42], image restoration [43], anisotropic diffusion [44], sparse-data surface representation [45], interpolation of missing image data [40, 46], and image binarization [34].

More recently, multigrid algorithms have been used to solve image reconstruction problems. Bouman and Sauer showed that nonlinear multigrid algorithms could be applied to inversion of Bayesian tomography problems [47]. This work used nonlinear multigrid techniques to compute maximum *a posteriori* (MAP) reconstructions with

non-Gaussian prior distributions and a non-negativity constraint. McCormick and Wade [48] applied multigrid methods to a linearized EIT problem, and Borcea [49] used a nonlinear multigrid approach to EIT based on a direct nonlinear formulation analogous to FAS in nonlinear multigrid PDE solvers. Brandt *et al.* developed multigrid methods for EIT [50] and atmospheric data assimilation [51], and applied multigrid or multiscale methods to various numerical computation problems including inverse problems [52, 53]. Johnson *et al.* [54] applied an algebraic multigrid algorithm to inverse bioelectric field problems formulated with the finite-element method. In [55, 56], Ye, *et al.* formulated the multigrid approach directly in an optimization framework, and used the method to solve ODT problems. In related work, Nash and Lewis formulated multigrid algorithms for the solution of a broad class of optimization problems [57, 58]. Importantly, both the approaches of Ye and Nash are based on the matching of cost functional derivatives at different scales.

In this paper, we propose a method we call multigrid inversion [59–62]. Multigrid inversion is a general approach for applying nonlinear multigrid optimization to the solution of inverse problems. A key innovation in our approach is that the resolution of both the forward and inverse models are varied. This makes our method particularly well suited to the solution of inverse problems with PDE forward models for a number of reasons:

- The computation can be dramatically reduced by using coarser grids to solve the forward model PDE. In previous approaches, the forward model PDE was solved only at the finest grid. This means that coarse grid updates were either computationally costly, or a linearization approximation was made for the coarse grid forward model [48, 55, 56].

- The coarse grid forward model can be modeled by a correctly discretized PDE, preserving the nonlinear characteristics of the forward model.

- A wide variety of optimization methods can be used for solving the inverse problem at each grid. Hence, common methods such as pre-conditioned con-

jugate gradient and/or adjoint differentiation [63, 64] can be employed at each grid resolution.

While the multigrid inversion method is motivated by the solution of inverse problems such as ODT and EIT, it is generally applicable to any inverse problem in which the forward model can be naturally represented at differing grid resolutions.

The multigrid inversion method is formulated in an optimization framework by defining a sequence of optimization functionals at decreasing resolutions. In order for the method to have well behaved convergence to the correct fine grid solution, it is essential that the cost functionals at different scales be consistent. To achieve this, we propose a recursive method for adapting the coarse grid functionals which guarantees that multigrid updates will not change an exact solution to the fine grid problem, i.e. that the exact fine grid solution is always a fixed point of the multigrid algorithm. In addition, we show that under certain conditions, the nonlinear multigrid inverse algorithm is guaranteed to produce monotone convergence of the fine grid cost functional. We present experimental results for the ODT application which show that the multigrid inversion algorithm can provide dramatic reductions in computation when the inversion problem is solved at the resolution necessary to achieve a high quality reconstruction.

This paper is organized as follows. Section 2.2 introduces the general concept of the multigrid inversion algorithm, and Section 2.2.4 discusses its convergence. In Section 2.3, we illustrate the application of the multigrid inversion method to the ODT problem, and its numerical results are provided in Section 2.4. Finally, Section 2.5 makes concluding remarks.

2.2 Multigrid Inversion Framework

In this section, we overview regularized inverse methods and then formulate the general multigrid inversion approach.

2.2.1 Inverse problems

Let Y be a random vector of (real or complex) measurements, and let x be a finite dimensional vector representing the unknown quantity, in our case an image, to be reconstructed. For any inverse problem, there is a forward model $f(x)$ given by

$$E[Y|x] = f(x) \tag{2.1}$$

which represents the computed means of the measurements given the image x. For many inverse problems, such as ODT, the forward model $f(x)$ is given by the solution of a PDE where x determines the coefficients of the discretized PDE. We will assume that the measurements Y are conditionally Gaussian given x, so that

$$\log p(y|x) = -\frac{1}{2\alpha}||y - f(x)||^2_\Lambda - \frac{P}{2}\log(2\pi\alpha|\Lambda|^{-1}) \ , \tag{2.2}$$

where Λ is a positive definite weight matrix, P is the dimensionality of the measurement, α is a parameter proportional to the noise variance, and $||w||^2_\Lambda = w^H\Lambda w$. Note that the measurement noise covariance matrix is equal to $\alpha\Lambda^{-1}$. When the data values are real valued, P is equal to the length of the vector Y, but when the measurements are complex, then P is equal to twice the dimension of Y.

Our objective is to invert the forward model of (2.1) and thereby estimate x from a particular measurement vector y. There are a variety of methods for performing this estimation, including maximum a posteriori (MAP) estimation, penalized maximum likelihood, and regularized inversion. All of these methods work by computing the value of x which minimizes a cost functional of the form

$$\frac{1}{2\alpha}||y - f(x)||^2_\Lambda + \frac{P}{2}\log(2\pi\alpha|\Lambda|^{-1}) + S(x) \ , \tag{2.3}$$

where $S(x)$ is a stabilizing functional used to regularize the inverse. Note that in the MAP approach, $S(x) = -\log p(x)$, where $p(x)$ is the prior distribution assumed for x. We will estimate both the noise variance parameter α and x by jointly maximizing over both quantities [65]. Minimization of (2.3) with respect to α yields the condition

$\hat{\alpha} = \frac{1}{P}||y - f(x)||^2_\Lambda$. Substitution of $\hat{\alpha}$ into (2.3) and dropping constants yields the cost functional to be optimized as

$$c(x) \quad = \quad \frac{P}{2} \log ||y - f(x)||^2_\Lambda + S(x) \ , \tag{2.4}$$

where we will generally assume $c(x)$ is a continuously differentiable function of x.

We have found that joint optimization over α and x has a number of important advantages. First, in many applications the absolute magnitude of the measurement noise is not known in advance, while the relative noise magnitude may be known. In such a scenario, it is useful to simultaneously estimate the value of α along with the value of x [55, 56, 66]. More importantly, we have found that the logarithm in the expression of (2.4) makes optimization less susceptible to being trapped in local minima. In any case, the multigrid methods we describe are equally applicable to the case when α is fixed. In this case, the cost functional is given by $c(x) = \frac{1}{2\alpha}||y - f(x)||^2_\Lambda + S(x)$, instead of (2.4).

2.2.2 Fixed-grid inversion

Once the cost functional of (2.4) is formulated, the inverse is computed by solving the associated optimization problem

$$\hat{x} = \arg\min_x \left\{ \frac{P}{2} \log ||y - f(x)||^2_\Lambda + S(x) \right\} \ . \tag{2.5}$$

Most optimization algorithms, such as CG, SD, and ICD, work by iteratively minimizing the cost functional. We express a single iteration of such a fixed grid optimizer as

$$x_{update} \leftarrow \text{Fixed_Grid_Update}(x_{init}, c(\cdot)) \ , \tag{2.6}$$

where $c(\cdot)$ is the cost functional being minimized, x_{init} is the initial value of x, and x_{update} is the updated value.[1] We will generally assume that the fixed grid

[1]We use the $\leftarrow$ symbol to denote assignment of a value to a variable, thereby eliminating the need for time indexing in update equations.

algorithm reduces the cost functional with each iteration, unless the initial value of x is at a local minimum of the cost functional. Therefore, we say that an update algorithm is *monotone* if $c(x_{update}) \leq c(x_{init})$, with strict inequality when $\nabla c(x_{init}) \neq 0$ or $x_{update} \neq x_{init}$. Repeated application of a monotone fixed grid optimizer will produce a sequence of estimates with monotonically decreasing cost. Thus, we may approximately solve (2.5) through iterative application of (2.6).

In many inverse problems, such as ODT, the forward model computation requires the solution of a 3-D PDE which must be discretized for numerical solution on a computer. Although a fine discretization grid is desirable because it reduces modeling error and increases the resolution of the final image, these improvements are obtained at the expense of a dramatic increase in computational cost. For a 3-D problem, the computational cost typically increases by a factor of 8 each time the resolution is doubled. Solving problems at fine resolution also tends to slow convergence. For example, many fixed grid algorithms such as ICD[2] effectively eliminate error at high spatial frequencies, but low frequency errors are damped slowly [10, 29].

2.2.3 Multigrid inversion algorithm

In this section, we derive the basic multigrid inversion algorithm for solving the optimization of (2.5). Let $x^{(0)}$ denote the finest grid image, and let $x^{(q)}$ be a coarse resolution representation of $x^{(0)}$ with a grid sampling period of 2^q times the finest grid sampling period. To obtain a coarser resolution image $x^{(q+1)}$ from a finer resolution image $x^{(q)}$, we use the relation $x^{(q+1)} = I_{(q)}^{(q+1)} x^{(q)}$, where $I_{(q)}^{(q+1)}$ is a linear decimation matrix. We use $I_{(q+1)}^{(q)}$ to denote the corresponding linear interpolation matrix.

We first define a coarse grid cost functional, $\tilde{c}^{(q)}(x^{(q)})$, with a form analogous to that of (2.4), but with quantities indexed by the scale q, as

$$\tilde{c}^{(q)}(x^{(q)}) = \frac{P}{2} \log ||y^{(q)} - f^{(q)}(x^{(q)})||_{\Lambda}^2 + S^{(q)}(x^{(q)}) \ . \tag{2.7}$$

[2]ICD is generally referred to as Gauss-Seidel in the PDE literature literature.

Notice that the forward model $f^{(q)}(\,\cdot\,)$ and the stabilizing functional $S^{(q)}(\,\cdot\,)$ are both evaluated at scale q. This is important because evaluation of the forward model at low resolution substantially reduces computation due to the reduced number of variables. The specific form of $f^{(q)}(\,\cdot\,)$ generally results from the physical problem being solved with an appropriate grid spacing. In Section 2.3, we will give a typical example for ODT where $f^{(q)}(\,\cdot\,)$ is computed by discretizing the 3-D PDE using a grid spacing proportional to 2^q. The quantity $y^{(q)}$ in (2.7) denotes an adjusted measurement vector at scale q. Note that in this work, we assume that $y^{(q)}$ and $f^{(q)}(\cdot)$ are of the same length at every scale q, so that the data resolution is not a function of q. The stabilizing functional at each scale is fixed and chosen to best approximate the fine scale functional. We give an example of such a stabilizing functional later in Section 2.2.5.

In the remainder of this section, we explain how the cost functionals at each scale can be matched to produce a consistent solution. To do this, we define an adjusted cost functional

$$\begin{aligned} c^{(q)}(x^{(q)}) &= \tilde{c}^{(q)}(x^{(q)}) - r^{(q)}x^{(q)} \\ &= \frac{P}{2}\log||y^{(q)} - f^{(q)}(x^{(q)})||^2_{\Lambda} + S^{(q)}(x^{(q)}) - r^{(q)}x^{(q)} \ , \end{aligned} \tag{2.8}$$

where $r^{(q)}$ is a row vector used to adjust the functional's gradient. At the finest scale, all quantities take on their fine scale values and $r^{(q)} = 0$, so that $\tilde{c}^{(0)}(x^{(0)}) = c^{(0)}(x^{(0)}) = c(x)$. Our objective is then to derive recursive expressions for the quantities $y^{(q)}$and $r^{(q)}$ that match the cost functionals at fine and coarse scales.

Let $x^{(q)}$ be the current solution at grid q. We would like to improve this solution by first performing an iteration of fixed grid optimization at the coarser grid $q+1$, and then using this result to correct the finer grid solution. This coarse grid update is

$$\tilde{x}^{(q+1)} \leftarrow \text{Fixed_Grid_Update}(I^{(q+1)}_{(q)}x^{(q)}, c^{(q+1)}(\cdot)) \ , \tag{2.9}$$

where $I_{(q)}^{(q+1)} x^{(q)}$ is the initial condition formed by decimating $x^{(q)}$, and $\tilde{x}^{(q+1)}$ is the updated value. We may now use this result to update the finer grid solution. We do this by interpolating the *change* in the coarser scale solution by

$$\tilde{x}^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)}) \ . \tag{2.10}$$

Ideally, the new solutions $\tilde{x}^{(q)}$ should be at least as good as the old solution $x^{(q)}$. Specifically, we would like $c^{(q)}(\tilde{x}^{(q)}) \leq c^{(q)}(x^{(q)})$ when the fixed grid algorithm is monotone. However, this may not be the case if the cost functionals are not consistent. In fact, for a naively chosen set of cost functionals, the coarse scale correction could easily move the solution away from the optimum.

This problem of inconsistent cost functionals is eliminated if the fine and coarse scale cost functionals are equal within an additive constant.[3] This means we would like

$$c^{(q+1)}(\tilde{x}^{(q+1)}) \cong c^{(q)}(x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)})) + \text{constant} \tag{2.11}$$

to hold for all values of $\tilde{x}^{(q+1)}$. Our objective is then to choose a coarse scale cost functional which matches the fine cost functional as described in (2.11). We do this by the proper selection of $y^{(q+1)}$ and $r^{(q+1)}$. First, we enforce the condition that the initial error between the forward model and measurements be the same at the coarse and fine scales, giving

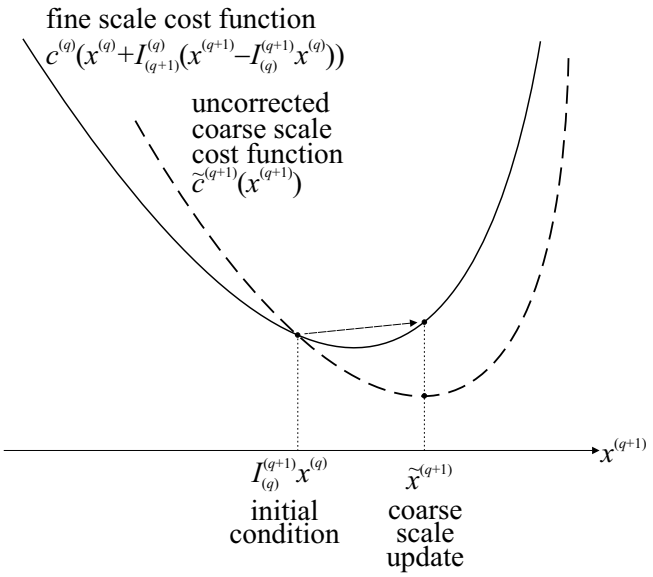
$$y^{(q+1)} - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) = y^{(q)} - f^{(q)}(x^{(q)}) \ . \tag{2.12}$$

This yields the update for $y^{(q+1)}$

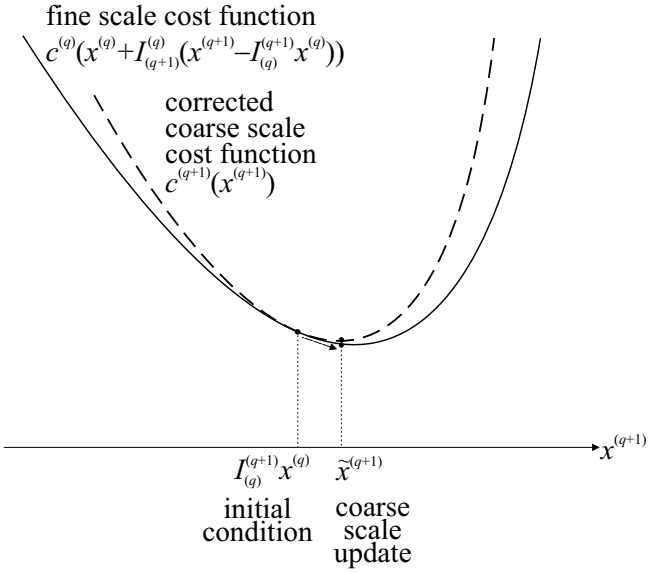
$$y^{(q+1)} \leftarrow y^{(q)} - \left[f^{(q)}(x^{(q)}) - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) \right] \ . \tag{2.13}$$

Intuitively, the term in the square brackets in (2.13) compensates for the forward model mismatch between resolutions.

[3] A constant offset has no effect on the value of x which minimizes the cost functional.



(a)



(b)

Fig. 2.1. The role of adjustment term $r^{(q+1)}x^{(q+1)}$. (a) When the gradients of the fine scale and coarse scale cost functionals are different at the initial value, the updated value may increase the fine grid cost functional's value. (b) When the gradients of the two functionals are matched, a properly chosen coarse scale functional can guarantee that the coarse scale update reduces the fine scale cost.

Next, we use the condition introduced in [55–58] to enforce the condition that the gradients of the coarse and fine cost functionals be equal at the current values of $x^{(q)}$ and $x^{(q+1)} = I_{(q)}^{(q+1)} x^{(q)}$. More precisely, we enforce the condition that

$$\nabla c^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} = \nabla c^{(q)}(x^{(q)}) I_{(q+1)}^{(q)} \ , \tag{2.14}$$

where $\nabla c(x)$ is the row vector formed by the gradient of the functional $c(\cdot)$. This condition is essential to assure that the optimum solution is a fixed point of the multigrid inversion algorithm [56], and is illustrated graphically in Fig. 2.1. In Section 2.2.4, we will also show how this condition can be used along with other assumptions to ensure monotone convergence of the multigrid inversion algorithm. Note that in (2.14), the interpolation matrix $I_{(q+1)}^{(q)}$, which comes from the chain rule of differentiation, actually functions like a decimation operator because it multiplies the gradient vector on the right. Importantly, the condition (2.14) holds for any choice of decimation and interpolation matrices.

The equality of (2.14) can be enforced at the current value $x^{(q)}$ by choosing

$$r^{(q+1)} \leftarrow \nabla \tilde{c}^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} - \left(\nabla \tilde{c}^{(q)}(x^{(q)}) - r^{(q)}\right) I_{(q+1)}^{(q)} \ , \tag{2.15}$$

where $\tilde{c}^{(q)}(\cdot)$ is the unadjusted cost functional defined in (2.7). By evaluating the gradients and using the update relation of (2.15), we obtain

$$r^{(q+1)} \leftarrow g^{(q+1)} - \left(g^{(q)} - r^{(q)}\right) I_{(q+1)}^{(q)} \ , \tag{2.16}$$

where $g^{(q)}$ and $g^{(q+1)}$ are the gradients of the unadjusted cost functional at the fine and coarse scales, respectively, given by

$$g^{(q)} = -\frac{P}{||y^{(q)} - f^{(q)}(x^{(q)})||_{\Lambda}^2} \mathrm{Re}\left\{\left(y^{(q)} - f^{(q)}(x^{(q)})\right)^H \Lambda A^{(q)}\right\} + \nabla S^{(q)}(x^{(q)}) \tag{2.17}$$

$$\begin{aligned} g^{(q+1)} = &-\frac{P}{||y^{(q)} - f^{(q)}(x^{(q)})||_{\Lambda}^2} \mathrm{Re}\left\{\left(y^{(q)} - f^{(q)}(x^{(q)})\right)^H \Lambda A^{(q+1)}\right\} \\ &+\nabla S^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}), \end{aligned} \tag{2.18}$$

where H is the conjugate transpose (Hermitian) operator, and $A^{(q)}$ denotes the gradient of the forward model or Fréchet derivative given by

$$A^{(q)} \ = \ \nabla f^{(q)}(x^{(q)}) \tag{2.19}$$

```
x^(q) ← Twogrid_Update(q, x^(q), y^(q), r^(q)) {
    Repeat ν_1^(q) times
        x^(q) ← Fixed_Grid_Update(x^(q), c^(q)(·; y^(q), r^(q))) //Fine grid update
    x^(q+1) ← I_(q)^(q+1) x^(q) //Decimation
    Compute y^(q+1) using (2.13)
    Compute r^(q+1) using (2.16)
    Repeat ν_1^(q+1) times
        x^(q+1) ← Fixed_Grid_Update(x^(q+1), c^(q+1)(·; y^(q+1), r^(q+1))) //Coarse grid update
    x^(q) ← x^(q) + I_(q+1)^(q)(x^(q+1) − I_(q)^(q+1) x^(q)) //Coarse grid correction
    Repeat ν_2^(q) times
        x^(q) ← Fixed_Grid_Update(x^(q), c^(q)(·; y^(q), r^(q))) //Fine grid update
    Return x^(q) //Return result
}
```

Fig. 2.2. Pseudo-code specification of a two-grid inversion algorithm. The notation $c^{(q+1)}(x^{(q+1)}; y^{(q+1)}, r^{(q+1)})$ is used to make the cost functional's dependency on $y^{(q+1)}$ and $r^{(q+1)}$ explicit.

$$A^{(q+1)} \quad = \quad \nabla f^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} \quad . \tag{2.20}$$

As a summary of this section, Fig. 2.2 shows pseudocode for implementing the two-grid algorithm. In this figure, we use the notation $c^{(q+1)}(x^{(q+1)}; y^{(q+1)}, r^{(q+1)})$ to make the dependency on $y^{(q+1)}$ and $r^{(q+1)}$ explicit. Notice that $\nu_1^{(q)}$ fixed grid iterations are done before the coarse grid correction, and that $\nu_2^{(q)}$ iterations are done afterwards. The convergence speed of the algorithm can be tuned through the choice of $\nu_1^{(q)}$ and $\nu_2^{(q)}$ at each scale.

The Multigrid-V algorithm [29] is obtained by simply replacing the fixed grid update at resolution $q+1$ of the two-grid algorithm with a recursive subroutine call, as shown in the pseudocode in Fig. 2.3(b). We can then solve (2.5) through iterative application of the Multigrid-V algorithm, as shown in Fig. 2.3(a). The Multigrid-V algorithm then moves from fine to coarse to fine resolutions with each iteration.

main( ) {
Initialize $x^{(0)}$ with a background estimate
$r^{(0)} \leftarrow 0$
$y^{(0)} \leftarrow y$
Choose number of fixed grid iterations $\nu_1^{(0)}, \ldots, \nu_1^{(Q-1)}$ and $\nu_2^{(0)}, \ldots, \nu_2^{(Q-1)}$
Repeat until converged:
$x^{(0)} \leftarrow \text{MultigridV}(q, x^{(0)}, c^{(0)}(\,\cdot\,; y^{(0)}, r^{(0)}))$
}

(a)

$x^{(q)} \leftarrow \text{MultigridV}(q, x^{(q)}, y^{(q)}, r^{(q)})$ {
Repeat $\nu_1^{(q)}$ times
$x^{(q)} \leftarrow \text{Fixed_Grid_Update}(x^{(q)}, c^{(q)}(\,\cdot\,; y^{(q)}, r^{(q)}))$ //Fine grid update
If $q = Q - 1$, return $x^{(q)}$ //If coarsest scale, return result
$x^{(q+1)} \leftarrow I_{(q)}^{(q+1)} x^{(q)}$ //Decimation
Compute $y^{(q+1)}$ using (2.13)
Compute $r^{(q+1)}$ using (2.15)
$x^{(q+1)} \leftarrow \text{MultigridV}(q+1, x^{(q+1)}, y^{(q+1)}, r^{(q+1)})$ //Coarse grid update
$x^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)}(x^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)})$ //Coarse grid correction
Repeat $\nu_2^{(q)}$ times
$x^{(q)} \leftarrow \text{Fixed_Grid_Update}(x^{(q)}, c^{(q)}(\,\cdot\,; y^{(q)}, r^{(q)}))$ //Fine grid update
Return $x^{(q)}$ //Return result
}

(b)

Fig. 2.3. Pseudo-code specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion. The Multigrid-V algorithm is similar to the 2-grid algorithm, but recursively calls itself to perform the coarse grid update.

2.2.4 Convergence of multigrid inversion

Multigrid inversion can be viewed as a method to simplify a potentially expensive optimization by temporarily replacing the original cost functional by a lower resolution one. In fact, there is a large class of optimization methods which depend on the use of so-called surrogate functionals, or functional substitution methods to speed or simplify optimization. A classic example of a surrogate functional is the Q-function used in the EM algorithm [67,68]. More recently, De Pierro discovered that this same basic method could be applied to tomography problems in a manner that allowed parallel updates of pixels in the computation of penalized ML reconstructions [69,70]. De Pierro's method has since been exploited to both prove convergence and allow parallel updates for ICD methods in tomography [71,72].

However, the application of surrogate functionals to multigrid inversion is unique in that the substituting functional is at a coarser scale and therefore has an argument of lower dimension. As with traditional approaches, the surrogate functional should be designed to guarantee monotone convergence of the original cost functional. In the case of the multigrid algorithm, a sequence of optimization functionals at varying resolutions should be designed so that the entire multigrid update decreases the finest resolution cost function.

Figure 2.1 graphically illustrates the use of surrogate functionals in multigrid inversion. Figure 2.1(a) shows the case in which the gradients of the fine scale and coarse scale (i.e. surrogate) functions are different at the initial value. In this case, the surrogate function can not upper bound the value of the fine scale functional, and the updated value may actually increase the fine grid cost functional's value. Figure 2.1(b) illustrates the case in which the gradients of the two functionals are matched. In this case, a properly chosen coarse scale functional can upper bound the fine scale functional, and the coarse scale update is guaranteed to reduce the fine scale cost.

The concepts illustrated in Fig. 2.1 can be formalized into conditions that guarantee the monotone convergence of the multigrid algorithms. The following theorem, proved in Appendix A, gives a set of sufficient conditions for monotone convergence of the multigrid inversion algorithm.

Theorem: (Multigrid Monotone Convergence)

For $0 \leq q < Q-1$, define the functional $\xi^{(q+1)} : I\!R^{N^{(q+1)}} \rightarrow I\!R$

$$\xi^{(q+1)}(x^{(q+1)}) = \tilde{c}^{(q+1)}(x^{(q+1)}) - \tilde{c}^{(q)}(x^{(q)} + I^{(q)}_{(q+1)}(x^{(q+1)} - I^{(q+1)}_{(q)}x^{(q)})) \ , \qquad (2.21)$$

where $N^{(q+1)}$ is the number of voxels in $x^{(q+1)}$, $I\!R$ is the set of real numbers, and the functions $\tilde{c}^{(q)}(\cdot)$ and $\tilde{c}^{(q+1)}(\cdot)$ are continuously differentiable. Assume that the following conditions are satisfied:

1. The fixed grid update is monotone for $0 \leq q < Q$.
2. $\xi^{(q)}(\,\cdot\,)$ is convex on $I\!R^{N^{(q)}}$ for $0 < q < Q$.
3. The adjustment vector $r^{(q+1)}$ is given by (2.15) for $0 \leq q < Q$.
4. $\nu_1^{(q)} + \nu_2^{(q)} \geq 1$ for $0 \leq q < Q$.

Then, the multigrid algorithm of Fig. 2.3 is monotone for $c^{(0)}(\,\cdot\,)$.

The conditions 1, 3, and 4 of the Theorem are easily satisfied for most problems. However, the difficulty lies in satisfying condition 2, convexity of $\xi^{(q)}(\cdot)$ for $q > 0$. If the eigenvalues of the Hessian of $\xi^{(q)}(\,\cdot\,)$ are lower-bounded, the convexity condition can be satisfied by adding a convex term, such as $\gamma||x^{(q)}||^2$, to $\tilde{c}^{(q)}(\,\cdot\,)$ for $q > 0$, where γ is a sufficiently large constant. However, addition of such a term tends to slow convergence by making the coarse scale corrections too conservative.

When the forward model is given by a PDE, it can be difficult or impossible to verify or guarantee the convexity condition of 2. Nonetheless, the theorem still gives insight into the convergence behavior of the algorithm; and in Section 2.4 we will show that empirically, for the difficult problem of ODT, the convergence of the multigrid algorithm is monotone in all cases, even without the addition of any convex terms.

2.2.5 Stabilizing functionals

The coarse scale stabilizing functionals, $S^{(q)}(x^{(q)})$, may be derived through appropriate scaling of $S(x)$. A general class of stabilizing functional has the form

$$S(x) = \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\,\rho\left(\frac{|x_i - x_j|}{\sigma}\right) , \tag{2.22}$$

where the set $\mathcal{N}$ consists of all pairs of adjacent grid points, b_{i-j} represents the weighting assigned to the pair $\{i,j\}$, σ is a parameter that controls the overall weighting, and $\rho(\cdot)$ is a symmetric function that penalizes the differences in adjacent pixel values. Such a stabilizing functional results from the selection of a prior density $p(x)$ corresponding to a Markov random field (MRF) [73]. A wide variety of functionals $\rho(\cdot)$ have been suggested for this purpose [74–76]. Generally, these methods attempt to select these functionals so that large differences in pixel value are not excessively penalized, thereby allowing the accurate formation of sharp edge discontinuities.

The stabilizing functional at scale q must be selected so that

$$S^{(q)}(x^{(q)}) \cong S(x) . \tag{2.23}$$

This can be done by using a form similar to (2.22) and applying scaling factors to result in

$$S^{(q)}(x^{(q)}) = 2^{qd} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\,\rho\left(\frac{|x_i^{(q)} - x_j^{(q)}|}{2^q\,\sigma}\right) , \tag{2.24}$$

where d is the dimension of the problem. Here we assume that $x_i - x_j \cong (x_i^{(q)} - x_j^{(q)})/2^q$, and we use the constant 2^{qd} to compensate for the reduction in the number of terms as the sampling grid is coarsened.

In our experiments, we use the generalized Gaussian Markov random field (GGMRF) image prior model [13, 14, 56, 76, 77] given by

$$p(x) = \frac{1}{\sigma^N z(p)} \exp\left[-\frac{1}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^p\right] , \tag{2.25}$$

where σ is a normalization parameter, $1 \leq p \leq 2$ controls the degree of edge smoothness, and $z(p)$ is a partition function. For the GGMRF prior, the stabilizing functional is given by

$$S(x) = \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i - x_j \right|^p \ , \tag{2.26}$$

and the corresponding coarse scale stabilizing functionals are derived using (2.24) to be

$$S^{(q)}(x^{(q)}) = \frac{1}{p(\sigma^{(q)})^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i^{(q)} - x_j^{(q)} \right|^p \ , \tag{2.27}$$

where $\sigma^{(q)}$ is given by

$$\sigma^{(q)} = 2^{q(1-\frac{d}{p})} \cdot \sigma^{(0)} \ . \tag{2.28}$$

2.3 Application to Optical Diffusion Tomography

Optical diffusion tomography is a method for determining spatial maps of optical absorption and scattering properties from measurements of light intensity transmitted through a highly scattering medium. In frequency domain ODT, the measured modulation envelope of the optical flux density is used to reconstruct the absorption coefficient and diffusion coefficient at each discretized grid point. However, for simplicity, we will only consider reconstruction of the absorption coefficient.

The complex amplitude $\phi_k(r)$ of the modulation envelope due to a point source at position s_k and angular frequency ω satisfies the frequency domain diffusion equation

$$\nabla \cdot [D(r) \nabla \phi_k(r)] + [-\mu_a(r) - j\omega/c] \phi_k(r) = -\delta(r - s_k) \ , \tag{2.29}$$

where r is position, c is the speed of light in the medium, $\mu_a(r)$ is the absorption coefficient, and $D(r)$ is the diffusion coefficient. The 3-D domain is discretized into N grid points, denoted by $r_1, r_2 \ldots, r_N$. The unknown image is then represented by an N dimensional column vector $x = [\mu_a(r_1), \mu_a(r_2), \ldots, \mu_a(r_N)]^T$ containing the absorption coefficients at each discrete grid point, where T is the transpose

operator. We will use the notation $\phi_k(r;x)$ in place of $\phi_k(r)$, in order to emphasize the dependence of the solution on the unknown image x. Then the measurement of a detector at location d_m resulting from a source at location s_k can be modeled by the complex value $\phi_k(d_m;x)$. The complete forward model function is then given by [4]

$$f(x) = [\ \phi_1(d_1;x), \phi_1(d_2;x), \ldots, \phi_1(d_M;x), \phi_2(d_1;x), \ldots, \phi_K(d_M;x)\ ]^T\ . \quad (2.30)$$

Note that $f(x)$ is a highly nonlinear function because it is given by the solution to a PDE using coefficients x. The measurement vector is also organized similarly as $y = [y_{11}, y_{12}, \ldots, y_{1m}, y_{21}, \ldots, y_{KM}]^T$, where y_{km} is the measurement with the source at s_k and the detector at d_m.

Our objective is to estimate the unknown image x from the measurements y. In a Bayesian framework, the MAP estimate of x is given by

$$\hat{x}_{MAP} = \arg\max_{x\geq 0}\{\ \log p(y|x) + \log p(x)\ \}\ , \quad (2.31)$$

where $p(y|x)$ is the data likelihood and $p(x)$ is the prior model for image x, which is assumed to be strictly positive in value. We use an independent Gaussian shot noise model (See [77] for details of this noise model) with the form given in (2.2), where the weight matrix Λ is given by

$$\Lambda = \mathrm{diag}(\frac{1}{|y_{11}|}, \ldots, \frac{1}{|y_{1M}|}, \frac{1}{|y_{21}|}, \ldots, \frac{1}{|y_{KM}|})\ . \quad (2.32)$$

For the prior model, we use the GGMRF density of (2.25) for $p(x)$. Using the formulation of Section 2.2.1, the ODT imaging problem is reduced to the optimization

$$\begin{aligned}(\hat{x}_{MAP}, \hat{\alpha}) \quad &= \quad \arg\max_{x\geq 0}\max_{\alpha} \\ &\left\{-\frac{1}{2\alpha}||y - f(x)||^2_\Lambda - \frac{P}{2}\log\alpha - \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^p\right\}\ ,\end{aligned} \quad (2.33)$$

[4]For simplicity of notation, we assume that all source-detector pairs are used. However, in our experimental simulations we use only a subset of all possible measurements. In fact, practical limitations can often limit the available measurements to a subset so that $P \neq 2KM$.

where constant terms are neglected. Minimizing (2.33) with respect to α reduces the cost functional to

$$c(x) = \frac{P}{2}\log||y - f(x)||_\Lambda^2 + \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^p \ . \tag{2.34}$$

This cost functional has the same form as (2.4) with the stabilizing functional given by (2.26). The gradient terms of the stabilizing functional used in (2.17) and (2.18) are given by

$$\nabla S(x) = \frac{1}{\sigma^p}\sum_{j\in\mathcal{N}_n} b_{n-j}|x_n - x_j|^{p-1}\mathrm{sgn}(x_n - x_j) \ . \tag{2.35}$$

We use multigrid inversion to solve the required optimization problem with coarse grid cost functionals of the form

$$\begin{aligned} c^{(q)}(x^{(q)}) &= \frac{P}{2}\log||y^{(q)} - f^{(q)}(x^{(q)})||_\Lambda^2 \\ &\quad + \frac{1}{p(\sigma^{(q)})^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|x_i^{(q)} - x_j^{(q)}\right|^p - r^{(q)}x^{(q)} \ , \end{aligned} \tag{2.36}$$

where $\sigma^{(q)}$ is given by (2.28) with $d = 3$.

At each scale q, we must also select a fixed grid optimization algorithm. For simplicity, we minimize (2.36) by alternatively minimizing with respect to α and x using the update formulas

$$\alpha \leftarrow \frac{1}{P}||y - f(x)||_\Lambda^2 \tag{2.37}$$

$$x \leftarrow\approx \arg\min_{x\geq 0}\left\{\frac{1}{2\alpha}||y - f(x)||_\Lambda^2 + \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|x_i - x_j\right|^p - rx\right\} \ , \tag{2.38}$$

where all expressions are interpreted as their corresponding scale q quantities. The fixed scale optimization (2.38) is performed using ICD optimization, as described in [77]. ICD requires the evaluation of the Fréchet derivative matrix of (2.19). For the ODT problem, it can be shown that the Fréchet derivative is given by [78]

$$\begin{aligned} A_{(k-1)M+m,\,n} &= \frac{\partial[f(x)]_{(k-1)M+m}}{\partial x_n} \\ &= \frac{\partial\phi_k(d_m; x)}{\partial x_n} \\ &= -G(s_k, r_n; x)G(d_m, r_n; x)V \ , \end{aligned} \tag{2.39}$$

where V is the voxel volume, $G(r_s, r_o; x)$ is the diffusion equation Green's function for the problem domain computed using the image x, with r_s as the source location and r_o as the observation point, and domain discretization errors are ignored [14,78]. Since the ODT problem is inherently 3-D, the Fréchet derivative matrix is usually very large. Fortunately, the separable structure of the Fréchet derivative can be use to substantially reduce memory requirements by storing the two quantities

$$\phi = [G(s_1, r_1; x), \ldots, G(s_1, r_N; x), G(s_2, r_1; x), \ldots, G(s_K, r_N; x)] \quad (2.40)$$

$$\psi = [G(d_1, r_1; x), \ldots, G(d_1, r_N; x), G(d_2, r_1; x), \ldots, G(d_M, r_N; x)] \quad (2.41)$$

and computing A on the fly [14].

The ICD algorithm is initialized by setting a state vector $\hat{y}$ equal to the forward model output for the current value of x, giving

$$\hat{y} \leftarrow f(x) \ . \quad (2.42)$$

Each ICD iteration is then computed by visiting each voxel n once using a random order, and updating each pixel value x_n and the state $\hat{y}$ using the following expressions

$$x_{old,n} \leftarrow x_n \quad (2.43)$$

$$x_n \leftarrow \arg\min_{u \geq 0} \left\{ \frac{1}{2\alpha} ||y - \hat{y} - A_{*n}(u - x_n)||^2_\Lambda + \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_n} b_{n-j} |u - x_j|^p - r_n u \right\} \quad (2.44)$$

$$\hat{y} \leftarrow \hat{y} + A_{*n}(x_n - x_{old,n}) \ , \quad (2.45)$$

where A_{*n} is the n^{th} column of the matrix A. Note that the state $\hat{y}$ keeps a running estimate of the forward model output by (2.45), so that subsequent state updates can be computed efficiently.

Figure 2.4 shows a detailed pseudo-code specification for the fixed grid and multi-grid algorithms for the ODT application. In particular, it explicitly shows the computation of the quantities $\phi^{(q)}$ and $\psi^{(q)}$ used in the computation of the Fréchet derivative.

```
main( ) {
    Initialize x^(0) with a background estimate
    For q = 1, 2, ..., Q − 1,      x^(q) ← I^(q)_(q−1) x^(q−1)
    For q = 0, 1, ..., Q − 1,      r^(q) ← 0    and    y^(q) ← y
    Repeat until converged: {
        Compute φ^(0), ψ^(0) and ŷ ← f^(0)(x^(0))
        If Multigrid Inversion :
            Choose ν_1^(0), ..., ν_1^(Q−1) and ν_2^(0), ..., ν_2^(Q−1)
            x^(0) ← MultigridV(0, x^(0), y^(0), r^(0), φ^(0), ψ^(0), ŷ)
        If Fixed Grid Inversion :
            x^(0) ← Fixed_Grid_Update(x^(0), y^(0), r^(0), φ^(0), ψ^(0), ŷ)
    }
}
```

(a)

$x \leftarrow \text{Fixed_Grid_Update}(x, y, r, \phi, \psi, \hat{y})$ {

Compute $\alpha \leftarrow \frac{1}{P}||y - \hat{y}||^2_\Lambda$

For $n = 0, \ldots, N-1$ (in random order), {

Compute column vector A_{*n} with (2.39)

Update x_n, as described by Ye, *et al.* [77]:

$x_{old,n} \leftarrow x_n$

$$x_n \leftarrow \arg\min_{u \geq 0} \left\{ \frac{1}{2\alpha} ||y - \hat{y} - A_{*n}(u - x_n)||^2_\Lambda + \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_n} b_{n-j}|u - x_j|^p - r_n u \right\}$$

$\hat{y} \leftarrow \hat{y} + A_{*n}(x_n - x_{old,n})$

}

}

(b)

$x^{(q)} \leftarrow \text{MultigridV}(q, x^{(q)}, y^{(q)}, r^{(q)}, \phi^{(q)}, \psi^{(q)}, \hat{y})$ {

For $\nu = 1, \ldots, \nu_1^{(q)}$

$x^{(q)} \leftarrow \text{Fixed_Grid_Update}(x^{(q)}, y^{(q)}, r^{(q)}, \phi^{(q)}, \psi^{(q)}, \hat{y})$ //Fine grid update

If $q = Q - 1$, return $x^{(q)}$ //If coarsest scale, return result

$x^{(q+1)} \leftarrow I^{(q+1)}_{(q)} x^{(q)}$ //Decimation

Compute $\phi^{(q+1)}$, $\psi^{(q+1)}$ and $\hat{y} \leftarrow f^{(q+1)}(x^{(q+1)})$

Compute $y^{(q+1)}$ using (2.13)

Compute $r^{(q+1)}$ using (2.16)

$x^{(q+1)} \leftarrow \text{MultigridV}(q+1, x^{(q+1)}, y^{(q+1)}, r^{(q+1)}, \phi^{(q+1)}, \psi^{(q+1)}, \hat{y})$ //Coarse grid update

$x^{(q)} \leftarrow x^{(q)} + I^{(q)}_{(q+1)}(x^{(q+1)} - I^{(q+1)}_{(q)} x^{(q)})$ //Coarse grid correction

For $\nu = 1, \ldots, \nu_2^{(q)}$

$x^{(q)} \leftarrow \text{Fixed_Grid_Update}(x^{(q)}, y^{(q)}, r^{(q)}, \phi^{(q)}, \psi^{(q)}, \hat{y})$ //Fine grid update

Return $x^{(q)}$ //Return result

}

(c)

Fig. 2.4. Pseudo-code specification of fixed grid and multigrid inversion methods for the ODT problem showing (a) main routine for ODT problems, (b) fixed-grid update, and (c) Multigrid-V inversion.

2.4 Numerical Results

This section contains the results of numerical experiments using simulated data sets. In all cases, our simulated physical measurements were generated using a $257 \times 257 \times 257$ grid discretization of the domain and the MUDPACK [79] PDE solver. We used the highest practical resolution for the forward model simulation, so as to achieve the best possible accuracy of the simulated measurements. Since the sources and detectors are not located exactly on the grid points, a three-dimensional linear interpolation of the nearest grid points was also used.

Our experiments used two tissue phantoms, which we refer to as the homogeneous and nonhomogeneous phantoms. Both phantoms had dimensions of $10 \times 10 \times 10$ cm, and each face contained eight sources and nine detectors with a single modulation frequency of 100 MHz, as shown in Fig. 2.5. So the number of sources was $K = 48$, and the the number of detectors was $M = 54$. Some experiments used all source/detector pairs ($P = 2KM = 5184$), while others only used source/detector pairs on different faces of the cube ($P = 2K(M/6) \times 5 = 4320$). A zero-flux boundary condition on the outer boundary was imposed to approximate the physical boundary condition [14, 77, 78].

The homogeneous phantom had the constant values $\mu_a = 0.02\ \text{cm}^{-1}$ and $D = 0.03$ cm. For the inhomogeneous phantom of Fig. 2.6(a), the μ_a background was linearly varied from $0.01\ \text{cm}^{-1}$ to $0.04\ \text{cm}^{-1}$ in a direction perpendicular to a surface of the cubic phantom, except for the outermost region of width 1.25 cm, which was homogeneous with $\mu_a = 0.025\ \text{cm}^{-1}$. Two spherical μ_a inhomogeneities with values of $\mu_a = 0.1\ \text{cm}^{-1}$ (left-top) and $\mu_a = 0.12\ \text{cm}^{-1}$ (right-bottom) were centered on the bisecting plane, which is parallel to the cubic phantom surfaces parallel to the background variation direction. The diffusion coefficient D was homogeneous with $D = 0.03$ cm. For both phantoms, the reconstruction was performed for all voxels except the eight, four, and two outermost layers of grid points for $65 \times 65 \times 65$, $33 \times 33 \times 33$, and $17 \times 17 \times 17$ reconstruction resolutions, respectively. These border

regions were fixed to their true values in order to avoid singularities near the sources and detectors. These regions have also been excluded from all cross-section figures and the evaluation of root-mean-square (RMS) reconstruction error.

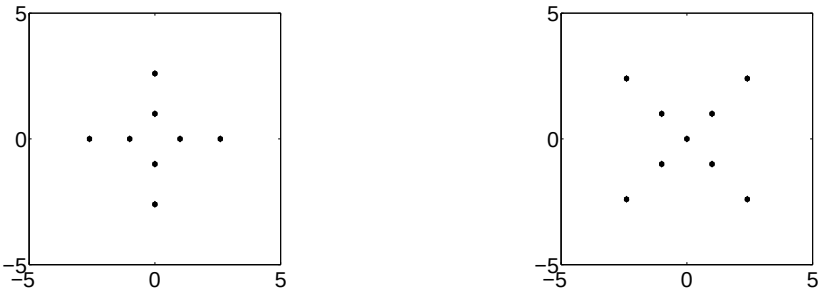
2.4.1 Evaluation of required forward model resolution

The objective of this section is to experimentally determine the forward model resolution required to produce a high quality reconstruction. To do this, we first evaluated the accuracy of the forward model as a function of resolution using the homogeneous phantom. The forward model PDE was first solved as resolutions corresponding to $129 \times 129 \times 129$, $65 \times 65 \times 65$, $33 \times 33 \times 33$, and $17 \times 17 \times 17$ grid points. We then computed the distortion-to-noise ratio (DNR) for two scenarios. The first scenario included all source/detector pairs, and the second only included source/detector pairs on different faces. This was done because the close proximity of source/detector pairs on the same face can result in susceptibility to discretization errors in the forward model. The DNR for the forward solution with l grid points on each side was computed as

$$DNR = \frac{2}{P} \sum_{i=1}^{P/2} \frac{\left| y_{(257)i} - y_{(l)i} \right|^2}{\left| y_{(257)i} \right|} , \tag{2.46}$$

where i is the index of source-detector pairs, $y_{(l)i}$ is the i-th forward solution with l grid points on each side, $y_{(257)i}$ is the i-th simulated measurement, which was computed with 257 grid points on each side, and $P/2$ is the number of complex measurements. Since $\left| y_{(257)i} \right|$ is proportional to the noise variance defined in (2.2) and (2.32), the DNR is proportional to the average ratio of discretization error and measurement noise.

Table 2.1 lists the DNR as a function of resolution for the two scenarios. Notice that for all resolutions the DNR is uniformly higher when source/detector pairs on the same face are included. As expected, the DNR also monotonically decreases as the resolution of the forward model is increased.



(a) (b)

Fig. 2.5. (a) Source and (b) detector pattern on each face of the cube geometry. Two data set scenarios were considered: one containing all source/detector pairs, and a second containing only source/detector pairs on different faces.

Table 2.1
Distortion-to-noise (DNR) ratio for various forward model resolutions. Coarse discretization increased forward model error, and source/detector pairs on the same face had much higher DNR.

Forward Model Resolution	Distortion-to-noise ratio	
	All measurements	Source/detector pairs on different faces
$17 \times 17 \times 17$	6.74×10^{-4}	9.96×10^{-7}
$33 \times 33 \times 33$	9.66×10^{-5}	2.85×10^{-8}
$65 \times 65 \times 65$	2.44×10^{-6}	3.35×10^{-9}
$129 \times 129 \times 129$	1.74×10^{-6}	1.04×10^{-10}

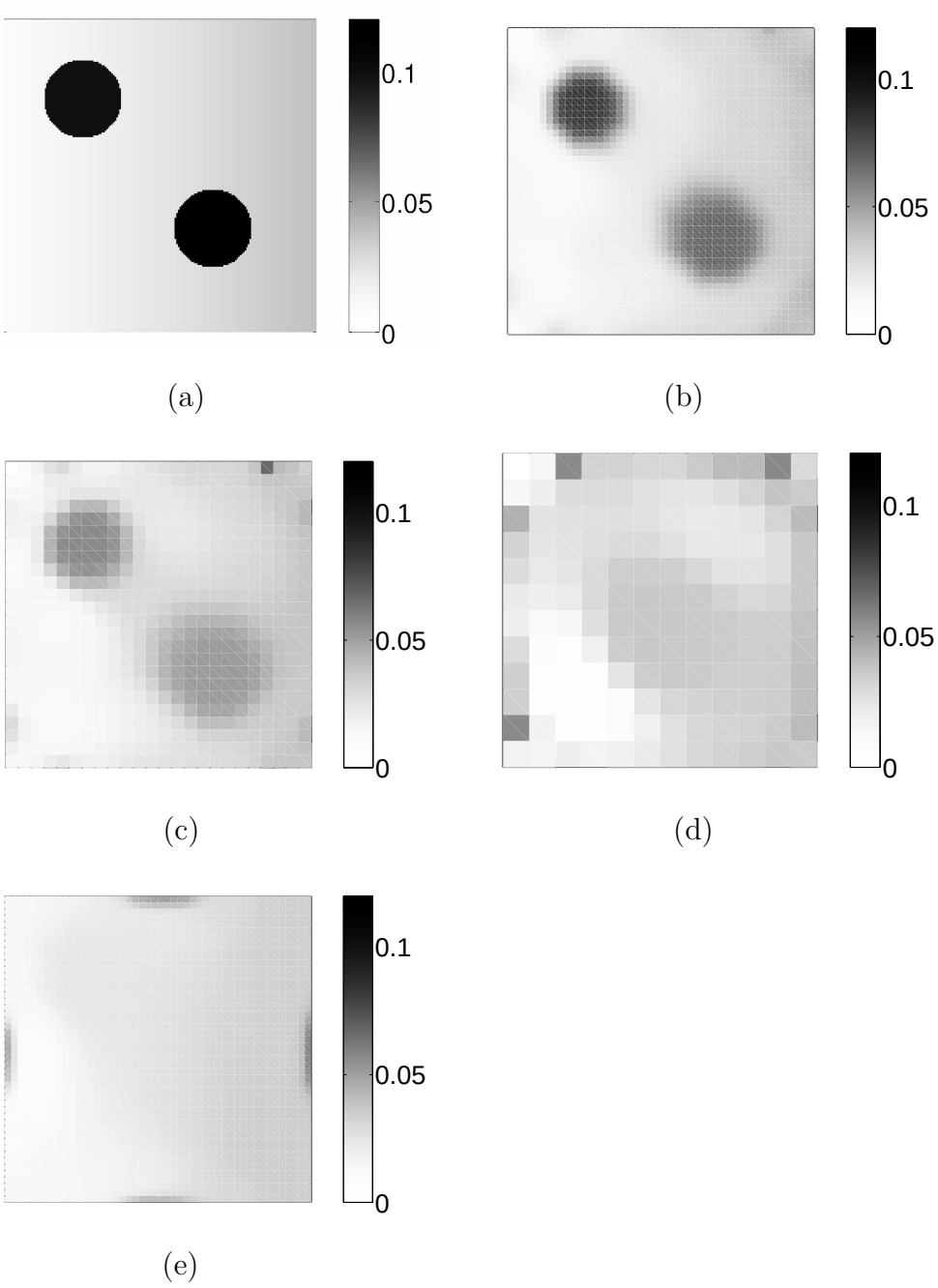


Fig. 2.6. A cross-section through (a) the inhomogeneous phantom, and the best reconstructions obtained using source detector pairs on different faces with (b) $65 \times 65 \times 65$ resolution, (c) $33 \times 33 \times 33$ resolution, (d) $17 \times 17 \times 17$ resolution, and (e) all source detector pairs with $65 \times 65 \times 65$ resolution.

Table 2.2
The normalization parameter σ that yields the best reconstruction and the resulting RMS image error between the reconstructions and the decimation of the true phantom.

Resolution/Data Set	σ	RMS image error
$65 \times 65 \times 65$/diff. faces	0.018	0.0069
$33 \times 33 \times 33$/diff. faces	0.008	0.0079
$17 \times 17 \times 17$/diff. faces	0.004	0.0093
$65 \times 65 \times 65$/all	0.03	0.0099

Next, we examined the reconstruction quality as a function of resolution using the inhomogeneous phantom. Gaussian shot noise was added to the data using Λ as given in (2.32) [77], so that the average signal-to-noise ratio for sources and detectors on opposite faces was 35 dB. Figure 2.6 shows a cross-section through the centers of inhomogeneities of the original phantom and the corresponding reconstructions for a variety of resolutions and data set scenarios.[5] Each reconstruction used $p = 1.2$, but the value of $\sigma = \sigma^{(0)}$ was chosen from in the range of 0.002 to 0.12, in order to minimize the RMS image error between the reconstructions and the decimation of the true phantom. The parameters and the resulting RMS errors are summarized in Table 2.2.

Figure 2.6 is consistent with the DNR measurement. The $65 \times 65 \times 65$ reconstruction from source/detector pairs on different faces has the best quality. Reconstructions at lower resolutions degrade rapidly, with very poor quality at $17 \times 17 \times 17$ resolution. Perhaps it is surprising that even the $65 \times 65 \times 65$ resolution reconstruction fails when all source/detector pairs are used. This result emphasizes the importance of using sufficiently high resolution, particularly when source/detector pairs are closely spaced.

2.4.2 Multigrid performance evaluation

The performance of the fixed-grid and multigrid algorithms was evaluated using the inhomogeneous phantom measurements of Sec. 2.4.1. Based on the results of Section 2.4.1, all comparisons of fixed-grid and multigrid inversion algorithms were performed for the $65 \times 65 \times 65$ resolution using only source/detector pairs on different faces. Our simulations compared fixed-grid inversion with multigrid inversion using 2, 3, and 4 levels of resolution. Table 2.3 lists these four cases together with our choice for the ν parameters at each scale. We selected the parameters ν to achieve

[5] These reconstructions were all produced using the multigrid algorithm with the mean phantom value as the initial condition because in each case this method converged to lowest cost among the tested algorithms.

robust convergence for a variety of problems. However, in other work [61], we have shown that these parameters can be adaptively chosen. The adaptive approach can further improve convergence speed and eliminates the need to select these parameters *a priori*. In order to make fair comparisons of computational speed, we scale the number of iterations for all methods into units of single fixed grid iterations at the finest scale. To do this, we use the approximate theoretical number of multiplies and the corresponding relative complexity shown in Table 2.3. However, we note that Table 2.3 indicates that the theoretical complexity of the multigrid iterations was somewhat lower then the experimentally measured complexity. See Appendix B for details of this conversion.

All reconstructions were done using the inhomogeneous phantom and a prior model with $p = 1.2$ and $\sigma = 0.018$ cm^{-1}. We chose $I_{(q)}^{(q+1)}$ to be the separable 3-D extensions of the 1-D decimation matrix

$$\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} \tag{2.47}$$

and $I_{(q+1)}^{(q)}$ to be the separable 3-D extension of the 1-D interpolation matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}, \tag{2.48}$$

respectively.

Table 2.3

Complexity comparison for each algorithm. Theoretical complex multiplications are estimated with (B.1) and theoretical relative complexity is the ratio of the required number of multiplications for one iteration to that for one fixed-grid iteration. Experimental relative complexity is the ratio of user time required for one iteration to that for one fixed-grid iteration.

Algorithm		Parameters							Theoretical		Experimental
		$\nu_1^{(0)}$	$\nu_2^{(0)}$	$\nu_1^{(1)}$	$\nu_2^{(1)}$	$\nu_1^{(2)}$	$\nu_2^{(2)}$	$\nu_1^{(3)}$	Multiplications ($\times 10^6$)	Relative Complexity	Relative Complexity
Fixed-grid		1	·	·	·	·	·	·	5,799	1	1
Multigrid-V	2 levels	1	·	20	·	·	·	·	24,569	4.23	4.96
	3 levels	1	·	5	5	40	·	·	21,479	3.70	4.56
	4 levels	1	·	4	4	20	20	60	20,775	3.58	4.60

For the first experiment, all algorithms were initialized with the average values of the true phantom, which were $\mu_a = 0.026$ cm^{-1} and $D = 0.03$ cm.[6] Figure 2.7 shows that the multigrid algorithms converged much faster than the fixed grid algorithm, both in the sense of cost and RMS error. The multigrid algorithms converged in only 20 iterations, while the fixed algorithm required 270 iterations. Even after 200 iterations, the fixed grid algorithm still changed very little in the convergence plots.

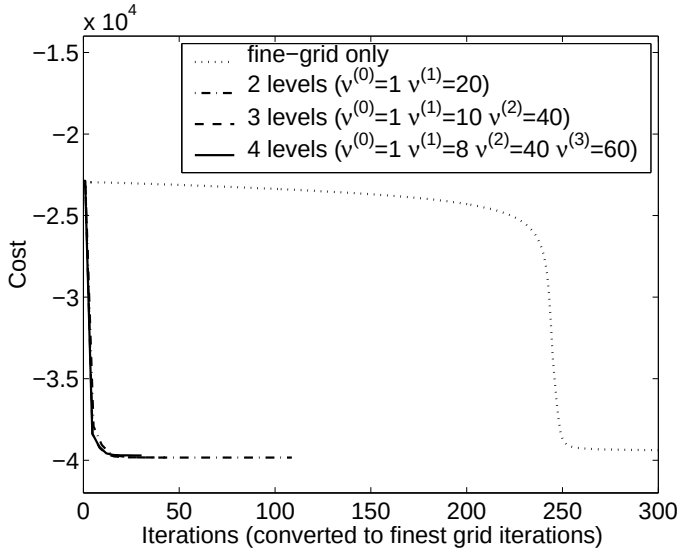
Figure 2.8 shows reconstructions produced by the four algorithms. The reconstructed image quality for all three multigrid algorithms is nearly identical, but the reconstructed quality is significantly worse for the fixed grid algorithm. In fact, the multigrid algorithms converged to slightly lower values of the cost functional (-3.9833×10^4 to -3.9763×10^4) than the fixed-grid algorithm (-3.9392×10^4), and the RMS image error for the multigrid reconstructions ranged from 0.0069 to 0.007, while the fixed algorithm converged to the higher RMS error of 0.0081.

To investigate the sensitivity of convergence with respect to initialization, we performed reconstructions with a poor initial estimate. The initial image was homogeneous, with a value of 1.75 times the true phantom's average value. The plots in Fig. 2.9 show that the three and four level multigrid algorithms converged rapidly. In particular, the four level multigrid algorithm converges almost as rapidly as it did when initialized with the true phantom's average value. The fixed grid algorithm changed very little from the initial estimate even after 300 iterations, and the two grid algorithm progressed slowly. These results suggest that higher level multigrid algorithms are necessary to overcome the effects of a poor initial estimate.

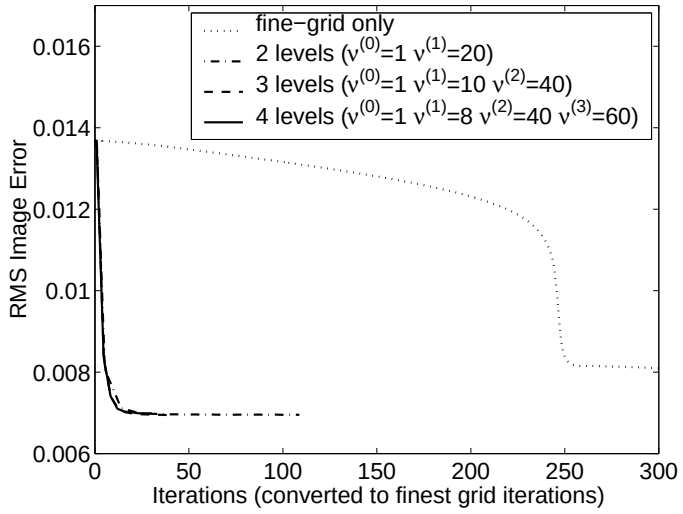
2.5 Conclusions

We have proposed a nonlinear multigrid inversion algorithm which works by simultaneously varying the resolution of both the forward model and inverse computation. Multigrid inversion is formulated in a general framework and is applicable to

[6] In practice, this is not possible since the average value is not known, but it was done because it favors the fixed-grid algorithm.



(a)



(b)

Fig. 2.7. Convergence of (a) cost function and (b) RMS image error when reconstructions were initialized with average values of true phantom. All multigrid algorithms converge about 13 times faster than the fixed-grid algorithm.

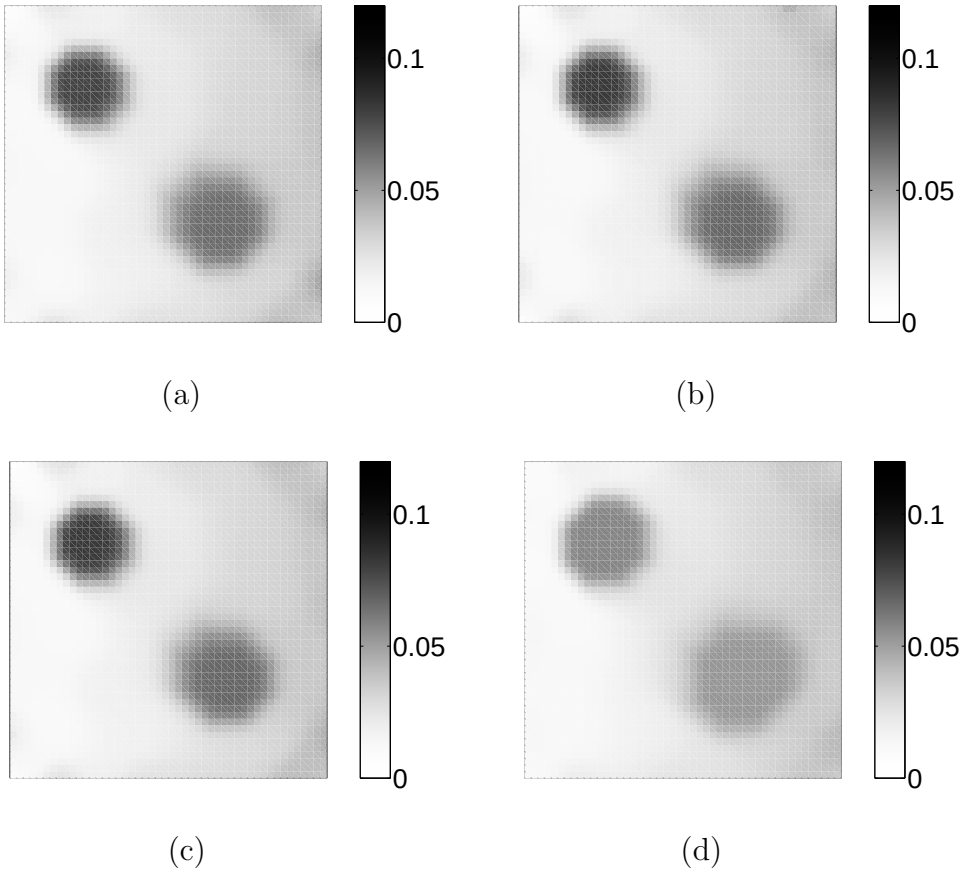
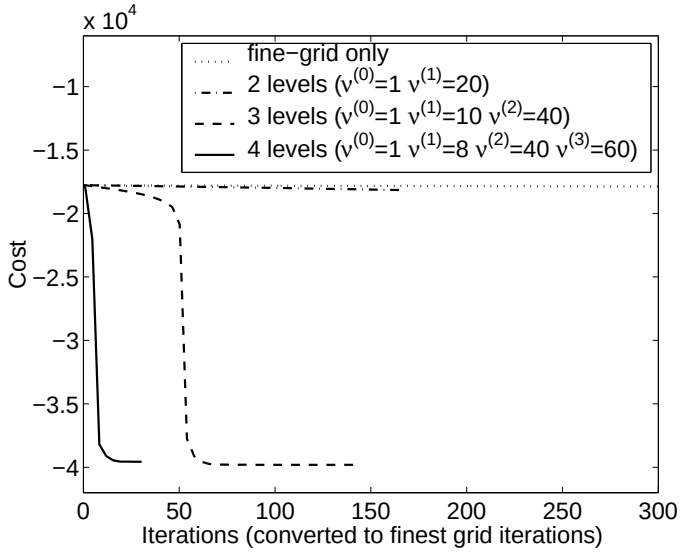
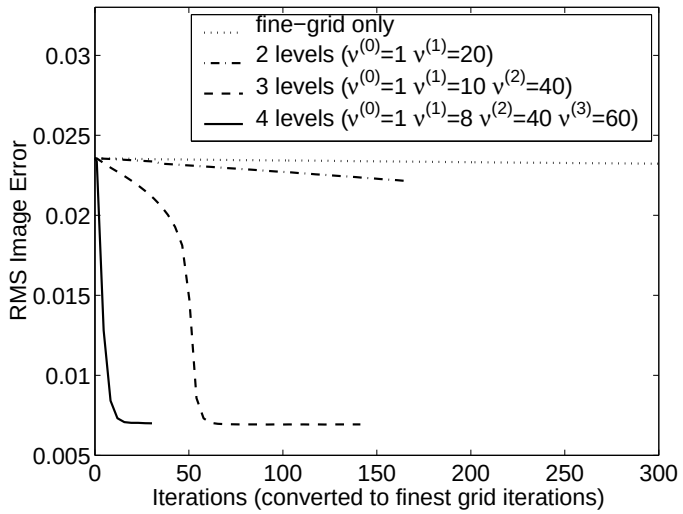


Fig. 2.8. Cross-sections of reconstructions on the plane through the centers of the inhomogeneities using (a) 4 level multigrid with 19.35 iterations, (b) 3 level multigrid with 19.95 iterations, (c) 2 level multigrid with 18.24 iterations, and (d) 270 fixed grid iterations. All the multigrid reconstructions have better image quality the the fixed grid reconstruction.



(a)



(b)

Fig. 2.9. Convergence of (a) cost function and (b) RMS image error with a poor initial guess. For higher level multigrid algorithms, the convergence was faster. In particular, the four level multigrid algorithm converged almost as fast as when the reconstruction was initialized with the true phantom's average value.

a wide variety of inverse problems, but it is particularly well suited for the inversion of nonlinear forward problems such as those modeled by the solution of PDEs.

We performed experimental simulations for the application of multigrid inversion to optical diffusion tomography using an ICD (Gauss-Seidel) fixed-grid optimizer. These simulations indicate that multigrid inversion can dramatically reduce computation, particularly if the reconstruction resolution is high, and the initial condition is inaccurate. Perhaps more importantly, multigrid inversion showed robust convergence under a variety of conditions and while solving an optimization problem that is subject to local minima. Future investigation could also make these comparisons using other fixed grid optimizers, such as conjugate gradient. Our experiments also indicated the importance of adequate resolution in the forward model.

3. MULTIGRID TOMOGRAPHIC INVERSION WITH VARIABLE RESOLUTION DATA AND IMAGE SPACES

3.1 Introduction

Over the past decade, many important image processing applications have been formulated in the framework of inverse problems. However, a major barrier to the use of inverse problem techniques has been the computational cost of these methods, which typically require the optimization of high dimensional and sometimes nonquadratic cost functionals. These computational challenges are only made more difficult by concurrent trends toward larger data sets and correspondingly higher resolution images in two and higher dimensions.

Multiresolution techniques have been widely investigated as a method for reducing the computation required to solve inverse problems. The techniques have ranged from simple coarse-to-fine approaches [11–15], which initialize fine scale iterations with coarse scale solutions, to more sophisticated wavelet or multiresolution image model-based approaches, which have been applied to image segmentation [80–83], image restoration [23, 84–88], and image reconstruction [16, 17, 20–26, 89].

Multigrid methods [27–29], which are multiresolution approaches originally developed for fast partial differential equation (PDE) solvers, have been recently applied to inverse problems such as image reconstruction [47, 48, 50–56, 90–92], optical flow estimation [33, 35–38], interpolation of missing image data [40, 46], image segmentation [40, 41], image analysis [33, 34, 42, 45], image restoration [43], and anisotropic diffusion [44]. Multigrid methods achieve fast convergence not only because coarse scale operations are much cheaper than those at fine scale, but also because coarse grid corrections typically remove low frequency error components more effectively

than fine scale corrections. Furthermore, unlike simple coarse-to-fine approaches, they provide a systematic method to go from fine to coarse, as well as from coarse to fine, so that coarse scale updates can be applied whenever they are expected to be effective. Since they operate directly in the space domain, multigrid algorithms can also easily enforce nonnegativity constraints, which are often necessary to obtain a physically meaningful image in tomographic reconstruction problems.

Interestingly, most of the existing work on multigrid image reconstruction has focused on applications that use a forward model described by the solution to one or more PDEs. For example, optical diffusion tomography (ODT) [55,56,91], electrical impedance tomography [48–50], bioelectric field problem [54], and atmospheric data assimilation [51] all use a forward model that depends implicitly on the solution to a PDE. In these applications multigrid algorithms provide significant computational savings, partly because good initialization is usually not available, and partly because per iteration computation tends to be high. For example, the application of our nonlinear multigrid inversion to ODT showed the potential for very large computational savings and robust convergence with respect to various operational initializations [91]. However, relatively little work has been done on applying multigrid methods to emission and transmission tomography problems [47,90,92].

Conventional tomography and many other inverse problems, such as motion analysis and image deblurring, have large measurement data sets which also can be decimated at coarse scales. Some inversion approaches have used multiresolution representations of this data. For example, wavelet decomposition of projection data is used in filtered backprojection [93–98] and MAP reconstruction [17,18,24], and a multiscale forward projection equation solver uses decimated sinogram data for coarse scale iterations [99]. Interestingly, the ordered subset expectation-maximization (OSEM) algorithm [100] does not use multiresolution data representation, but it does use only a subset of the data in each iteration. Importantly, existing multigrid methods, including our previous multigrid inversion framework [91], do not exploit the possibility of coarse representation of measurement data at coarser scales, and

thus their computational gain comes only from a reduced number of unknown variables by coarse discretization of the image at coarser scales.

In this paper, we propose a new multigrid method that is novel in three important ways. First, it reduces computation by changing the resolution of the data space as well as the image space. Second, it formulates the multigrid inversion problem for Bayesian reconstruction from transmission or emission data with either a Poisson or Gaussian noise model. Third, it incorporates a novel adaptive multigrid scheme which allocates computation to the scale at which the algorithm can best reduce the cost [61].

As with our previous multigrid inversion method [91], our new multigrid method formulates a consistent set of coarse scale cost functions and moves up and down recursively in resolution to solve the original finest scale problem. However, the important difference from our previous formulation is that the measurement data as well as the image is coarsely discretized at the coarse scale, and thus computation is further reduced. This is especially advantageous in applications where the data as well as the image have high dimension.

An important feature of our formulation is that the choice of decimator/interpolator for the data space is independent of the choice of those for the image space. In many image processing applications, such as motion analysis and image deblurring, a measurement is available for each pixel of the image space, so the same decimation/interpolation operators may be using on both the data and images. However, in many applications, including tomography, this is not true. Thus, the flexibility in choosing the decimator/interpolator makes our proposed multigrid approach particulary suitable for tomographic image reconstruction problems.

Our simulation results show that our multigrid algorithms using variable data resolution yield better convergence speed than the iterative coordinate descent (ICD) method [10, 101] and multigrid algorithms using fixed data resolution.

3.2 Multigrid Inversion with Variable Resolution Data and Image Spaces

In this section we present a multigrid inversion approach that changes resolutions of both data and image spaces. We first present our approach for the case of measurements with additive Gaussian noise, and we then generalize the method for inversion with Poisson noise.

3.2.1 Quadratic data term case

Let $Y \in I\!R^M$ be a random vector of measured data, and let $x \in I\!R^N$ be a discretized unknown image. Then, the expected value of the measurement vector is given by

$$E[Y|x] = f(x) \tag{3.1}$$

where $f : I\!R^N \rightarrow I\!R^M$ is know as the forward model. Our task is then to estimate the image x which produced the observations Y. A common approach for solving this problem is to solve an associated optimization problem of the form

$$\hat{x} = \arg\min_x \{-\log p(y|x) + S(x)\} \ , \tag{3.2}$$

where $p(y|x)$ is the probability density of Y given x, and $S(x)$ is a stabilizing function designed to regularize the inversion [102, 103]. If $S(x) = -\log p(x)$, where $p(x)$ is the image prior probability density, this results in the maximum *a posteriori* (MAP) estimate of x.

If the measurements Y are conditionally Gaussian given x with noise covariance matrix $(2\Lambda)^{-1}$, then the inverse is computed by minimizing the cost function

$$||y - f(x)||^2_\Lambda + S(x) \ , \tag{3.3}$$

where $||w||^2_\Lambda = w^H \Lambda w$. By expanding the data term of (3.3), the cost function may be expressed within a constant as

$$c(x) \ = \ ||f(x)||^2_\Lambda + 2a^T f(x) + S(x) \ , \tag{3.4}$$

where $a = -\Lambda^T y$. For the case where we estimate a noise scaling parameter, see Appendix D.

Minimizing a function such as (3.4) can be very computationally expensive, particularly when the image x and data y have high dimension. Our approach to reducing computation will be to formulate an approximate cost function using a coarse scale representation of the image and data. To do this, we require methods for decimating and interpolating in both domains.

Let $x^{(q)} \in I\!R^{N^{(q)}}$ and $y^{(q)} \in I\!R^{M^{(q)}}$ denote representations of $x = x^{(0)}$ and $y = y^{(0)}$ at coarser resolution q. In order to convert between resolutions, we define the image domain decimation operator $x^{(q+1)} = I_{(q)}^{(q+1)} x^{(q)}$ and the data domain decimation operator $y^{(q+1)} = J_{(q)}^{(q+1)} y^{(q)}$. Similarly, we define the interpolation operators for image and data domains as $x^{(q)} = I_{(q+1)}^{(q)} x^{(q+1)}$ and $y^{(q)} = J_{(q+1)}^{(q)} y^{(q+1)}$, respectively. Typically, we use either pixel replication or bilinear interpolation operators and decimation operators, but the theory is applicable to a wide range of choices. Notice that in general, $I_{(q)}^{(q+1)}$ and $J_{(q)}^{(q+1)}$ may be different.

We will assume that there is some natural way to define a coarse scale forward model $f^{(q)} : I\!R^{N^{(q)}} \rightarrow I\!R^{M^{(q)}}$ which maps the coarse scale image to the coarse scale data. In practice, $f^{(q)}(\cdot)$ can result from the method used to discretize the physical problem, but at this point we will make few assumptions regarding its specific form. The most crucial assumption in our formulation is that

$$f^{(0)}(x^{(0)}) \;\; \cong \;\; J_{(q)}^{(0)} f^{(q)}(x^{(q)}) \; . \tag{3.5}$$

Then by replacing $f^{(0)}(x^{(0)})$ in the original finest scale cost function (3.4) with an interpolated forward model $J_{(q)}^{(0)} f^{(q)}(x^{(q)})$, we have an approximate coarse scale cost function

$$\tilde{c}^{(q)}(x^{(q)}) \;\; = \;\; ||J_{(q)}^{(0)} f^{(q)}(x^{(q)})||_{\Lambda}^2 + 2a^T J_{(q)}^{(0)} f^{(q)}(x^{(q)}) + S^{(q)}(x^{(q)}) \; , \tag{3.6}$$

where the coarse scale stabilizing function $S^{(q)}(\cdot)$ is chosen to best approximate the original finest scale one, as described in [91] and later in Sec. 3.4.1. By defining

$$\Lambda^{(q)} \;\; = \;\; [J_{(q)}^{(0)}]^T \Lambda^{(0)} J_{(q)}^{(0)} \tag{3.7}$$

$$a^{(q)} \;=\; [J_{(q)}^{(0)}]^T a^{(0)} \;, \tag{3.8}$$

(3.6) can be expressed as

$$\tilde{c}^{(q)}(x^{(q)}) \;=\; ||f^{(q)}(x^{(q)})||^2_{\Lambda^{(q)}} + 2a^{(q)^T} f^{(q)}(x^{(q)}) + S^{(q)}(x^{(q)}) \;. \tag{3.9}$$

The form of (3.9) is analogous to that of (3.4), but with quantities indexed by the scale q. As in our previous work [91], the forward model $f^{(q)}(\,\cdot\,)$ and the stabilizing function $S^{(q)}(\,\cdot\,)$ use a coarsely discretized image at each scale q, and thus computations are substantially reduced due to the reduced number of variables. In this work, computation is further reduced since the dimension of the forward model vector also changes with q.

We adjust the coarse scale cost functions (3.9) at each scale to better match with the original fine scale one, and thus to produce a consistent solution. To do this, we define an adjusted cost function by appending an additional linear correction term. This yields the adjusted cost function

$$c^{(q)}(x^{(q)}) \;=\; ||f^{(q)}(x^{(q)})||^2_{\Lambda^{(q)}} + 2a^{(q)^T} f^{(q)}(x^{(q)}) + S^{(q)}(x^{(q)}) - r^{(q)}x^{(q)} \;, \tag{3.10}$$

where $r^{(q)}$ is a row vector used to adjust the function's gradient, the choice of which will be discussed later. At the finest scale, $r^{(q)} = 0$ is chosen so that $c^{(0)}(x^{(0)}) = c(x)$.

With the set of coarse scale cost functions of the form in (3.10), the multigrid algorithm solves the original problem by moving up and down in resolution [56, 91]. Let $x^{(q)}$ be the current solution at grid q. We would like to improve this solution by first performing iterations of fixed grid optimization at the coarser grid $q+1$, and then using this result to correct the finer grid solution. This coarse grid update is

$$\tilde{x}^{(q+1)} \leftarrow \text{Fixed_Grid_Update}(I_{(q)}^{(q+1)} x^{(q)}, c^{(q+1)}(\cdot)) \;, \tag{3.11}$$

where $\tilde{x}^{(q+1)}$ is the updated value, and the operator Fixed_Grid_Update$(x_{init}, c(\cdot))$ is any fixed grid update algorithm designed to reduce the cost function $c(\cdot)$ starting with the initial value x_{init}. In (3.11), the initial condition $I_{(q)}^{(q+1)} x^{(q)}$ is formed by

decimating $x^{(q)}$. We may now use this result to update the finer grid solution. We do this by interpolating the *change* in the coarser scale solution.

$$\tilde{x}^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}) \ . \tag{3.12}$$

In order to ensure updates which reduce the fine scale cost, we would like to make the fine and coarse scale cost functions equal within an additive constant. This means we would like the equation

$$c^{(q+1)}(\tilde{x}^{(q+1)}) \cong c^{(q)}\left(x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)})\right) + \text{constant} \tag{3.13}$$

to hold for all coarse-scale updated values of $\tilde{x}^{(q+1)}$. Our objective is then to choose a coarse scale cost function which matches the fine cost function, as described in (3.13). We do this by selecting $r^{(q+1)}$ to match the gradients of the coarse and fine cost functions at the current values of $x^{(q)}$ and $x^{(q+1)} = I_{(q)}^{(q+1)}x^{(q)}$. More precisely, we enforce the condition that

$$\nabla c^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} = \nabla c^{(q)}(x^{(q)})I_{(q+1)}^{(q)} \ , \tag{3.14}$$

where $\nabla c(x)$ is the row vector formed by the gradient of the function $c(\cdot)$ [56]. This condition (3.14) is essential to assure that the optimum solution is a fixed point of the multigrid inversion algorithm [56], and we can show how this condition can be used along with other assumptions to ensure monotone convergence of the multigrid inversion algorithm [91]. Note that in (3.14), the interpolation matrix $I_{(q+1)}^{(q)}$, which comes from the chain rule of differentiation, actually functions like a decimation operator because it multiplies the gradient vector on the right. Importantly, the condition (3.14) holds for any choice of decimation and interpolation matrices. The equality of (3.14) can be enforced at the current value $x^{(q)}$ by choosing

$$r^{(q+1)} \leftarrow \nabla\tilde{c}^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} - \left(\nabla\tilde{c}^{(q)}(x^{(q)}) - r^{(q)}\right) I_{(q+1)}^{(q)} \ . \tag{3.15}$$

By evaluating the gradient for the cost function (3.4), (3.15) is computed by

$$r^{(q+1)} \leftarrow g^{(q+1)} - \left(g^{(q)} - r^{(q)}\right) I_{(q+1)}^{(q)} \ , \tag{3.16}$$

where $g^{(q)}$ and $g^{(q+1)}$ are the gradients of the unadjusted cost function at the fine and coarse scales, respectively, given by

$$g^{(q)} \leftarrow 2\left(f^{(q)}(x^{(q)})\Lambda^{(q)^T} + a^{(q)}\right)^T A^{(q)} + \nabla S^{(q)}(x^{(q)}) \tag{3.17}$$

$$g^{(q+1)} \leftarrow 2\left(f^{(q+1)}(x^{(q+1)})\Lambda^{(q+1)^T} + a^{(q+1)}\right)^T A^{(q+1)} + \nabla S^{(q+1)}(x^{(q+1)}) \ , \tag{3.18}$$

where T is the transpose operator, and $A^{(q)}$ denotes the gradient of the forward model or Fréchet derivative given by

$$A^{(q)} = \nabla f^{(q)}(x^{(q)}) \tag{3.19}$$

$$A^{(q+1)} = \nabla f^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} \ . \tag{3.20}$$

Assuming that

$$J_{(q+1)}^{(0)} = J_{(q)}^{(0)} J_{(q+1)}^{(q)} \ , \tag{3.21}$$

the coarse scale cost function parameters (3.7) (3.8) can be computed iteratively by

$$\Lambda^{(q+1)} \leftarrow [J_{(q+1)}^{(q)}]^T \Lambda^{(q)} J_{(q+1)}^{(q)} \tag{3.22}$$

$$a^{(q+1)} \leftarrow [J_{(q+1)}^{(q)}]^T a^{(q)} \ . \tag{3.23}$$

The computations of (3.22) and (3.23) are inexpensive and, in addition, can be precomputed since they are independent of the image $x^{(q)}$.

The pseudocode in Fig. 3.1(b) shows the Multigrid-V algorithm to solve the minimization of (3.4). Multigrid-V recursion is a standard multigrid methods, which calls itself recursively in resolution. More specifically, it replaces the coarse scale fixed-grid update of (3.11) by a recursive call of multigrid algorithm. We solve the problem through iterative application of the Multigrid-V algorithm, as shown in Fig. 3.1(a). See [27–29, 56, 91] for the details of Multigrid-V recursion.

3.2.2 Poisson data case

Some inverse problems, such as transmission and emission tomography, use Poisson measurement noise models [104, 105]. In the Poisson noise model, we assume

```
main( ) {
    Initialize x^(0) with a background estimate
    For q = 0, 1, ..., Q − 2,      x^(q+1) ← I_(q)^(q+1) x^(q)
    For q = 0, 1, ..., Q − 1,      r^(q) ← 0
    If Gaussian noise model is used, then {
        For q = 0, 1, ..., Q − 2,      Λ^(q+1) ← [J_(q+1)^(q)]^T Λ^(q) J_(q+1)^(q)
        For q = 0, 1, ..., Q − 2,      a^(q+1) ← [J_(q+1)^(q)]^T a^(q)
    }
    If Poisson noise model is used, then {
        For q = 1, 2, ..., Q − 1,      y^(q) ← J_(0)^(q) y^(0)
    }
    Choose number of fixed grid iterations ν_1^(0), ..., ν_1^(Q−1) and ν_2^(0), ..., ν_2^(Q−1)
    Repeat until converged:
        x^(0) ← MultigridV(q, x^(0), r^(0))
}
```

(a)

```
x^(q) ← MultigridV(q, x^(q), r^(q)) {
    Repeat ν_1^(q) times
        x^(q) ← Fixed_Grid_Update(x^(q), c^(q)(·; r^(q))) //Fine grid update
    If q = Q − 1, return x^(q) //If coarsest scale, return result
    x^(q+1) ← I_(q)^(q+1) x^(q) //Decimation
    If Gaussian noise model is used, then {
        Compute r^(q+1) using (3.15) (3.17) and (3.18)
    }
    If Poisson noise model is used, then {
        Compute r^(q+1) using (3.15) (3.33) and (3.34)
    }
    x^(q+1) ← MultigridV(q + 1, x^(q+1), r^(q+1)) //Coarse grid update
    x^(q) ← x^(q) + I_(q+1)^(q) (x^(q+1) − I_(q)^(q+1) x^(q)) //Coarse grid correction
    Repeat ν_2^(q) times
        x^(q) ← Fixed_Grid_Update(x^(q), c^(q)(·; r^(q))) //Fine grid update
    Return x^(q) //Return result
}
```

(b)

Fig. 3.1. Pseudo-code specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion.

(3.1) holds with the Y_m's being independent Poisson random variables. In this case, the negative log likelihood of the Poisson data is given by

$$-\log p(y|x) = \sum_{m=1}^{M} \{f_m(x) - y_m \log f_m(x) + \log(y_m!)\} \ , \tag{3.24}$$

where M is the number of measurements and y_m is a realization of Y_m, and its corresponding regularized inverse can be solved by minimizing the cost function

$$c(x) = \sum_{m=1}^{M} \{f_m(x) - y_m \log f_m(x)\} + S(x) \ . \tag{3.25}$$

We first compute coarse scale measurement data using data domain decimation

$$y^{(q)} \stackrel{\triangle}{=} J_{(0)}^{(q)} y^{(0)} \ . \tag{3.26}$$

In addition to (3.5), we also make a few assumptions, which are satisfied for most choices of data domain decimation and interpolation operators. First, we assume that the interpolated coarse scale data approximates the fine scale data. More formally, we say

$$y^{(0)} \stackrel{\simeq}{=} J_{(q)}^{(0)} y^{(q)} \ . \tag{3.27}$$

Second, we assume that

$$f_m^{(0)}(x^{(0)}) \stackrel{\simeq}{=} f_i^{(q)}(x^{(q)}) \quad \text{for } \left[J_{(q)}^{(0)}\right]_{m,i} \neq 0 \ , \tag{3.28}$$

where $[B]_{m,i}$ is the $(m,i)^{th}$ element of matrix B. In order to understand this assumption, notice that when $\left[J_{(q)}^{(0)}\right]_{m,i}$ is nonzero, m and i are the corresponding data at different resolutions. So in this case, we would expect the two data to be approximately equal. Third, we assume that

$$\sum_{m=1}^{M^{(0)}} \left[J_{(q)}^{(0)}\right]_{m,i} = \frac{M^{(0)}}{M^{(q)}} \ , \tag{3.29}$$

which insures that the average value of $y^{(0)}$ and $y^{(q)}$ are the same.

The negative logarithm of the Poisson data likelihood (3.24) can then be approximated as

$$\begin{aligned}
&-\log p(y|x) - \sum_{m=1}^{M} \log(y_m!) \\
&= \sum_{m=1}^{M^{(0)}} \left\{ f_m^{(0)}(x^{(0)}) - y_m^{(0)} \log f_m^{(0)}(x^{(0)}) \right\} \\
&\simeq \sum_{m=1}^{M^{(0)}} \left\{ \left[J_{(q)}^{(0)} f^{(q)}(x^{(q)}) \right]_m - \left[J_{(q)}^{(0)} y^{(q)} \right]_m \log f_m^{(0)}(x^{(0)}) \right\} \\
&= \sum_{m=1}^{M^{(0)}} \left\{ \sum_{i=1}^{M^{(q)}} \left[J_{(q)}^{(0)} \right]_{m,i} f_i^{(q)}(x^{(q)}) - \sum_{i=1}^{M^{(q)}} \left[J_{(q)}^{(0)} \right]_{m,i} y_i^{(q)} \log f_m^{(0)}(x^{(0)}) \right\} \\
&\simeq \sum_{m=1}^{M^{(0)}} \left\{ \sum_{i=1}^{M^{(q)}} \left[J_{(q)}^{(0)} \right]_{m,i} f_i^{(q)}(x^{(q)}) - \sum_{i=1}^{M^{(q)}} \left[J_{(q)}^{(0)} \right]_{m,i} y_i^{(q)} \log f_i^{(q)}(x^{(q)}) \right\} \\
&= \sum_{i=1}^{M^{(q)}} \left\{ \left(f_i^{(q)}(x^{(q)}) - y_i^{(q)} \log f_i^{(q)}(x^{(q)}) \right) \sum_{m=1}^{M^{(0)}} \left[J_{(q)}^{(0)} \right]_{m,i} \right\} \\
&= \frac{M^{(0)}}{M^{(q)}} \sum_{i=1}^{M^{(q)}} \left[f_i^{(q)}(x^{(q)}) - y_i^{(q)} \log f_i^{(q)}(x^{(q)}) \right] , \qquad (3.30)
\end{aligned}$$

where the third line comes from (3.5) and (3.27), the fourth from the element-by-element expansion of the data domain interpolation, the fifth from (3.28), the sixth from the summation order exchange, and the last from (3.29). Thus, an approximate coarse scale cost function with a reduced resolution data and forward model may be expressed as

$$\tilde{c}^{(q)}(x^{(q)}) = \frac{M^{(0)}}{M^{(q)}} \sum_{m=1}^{M^{(q)}} \left[f_m^{(q)}(x^{(q)}) - y_m^{(q)} \log f_m^{(q)}(x^{(q)}) \right] + S^{(q)}(x^{(q)}) \ . \qquad (3.31)$$

The adjusted coarse scale cost is then obtained by adding the gradient correction term

$$c^{(q)}(x^{(q)}) = \frac{M^{(0)}}{M^{(q)}} \sum_{m=1}^{M^{(q)}} \left\{ f_m^{(q)}(x^{(q)}) - y_m^{(q)} \log f_m^{(q)}(x^{(q)}) \right\} + S^{(q)}(x^{(q)}) - r^{(q)} x^{(q)} \ , \qquad (3.32)$$

where $r^{(q)}$ is computed by (3.16) with

$$g^{(q)} \leftarrow \frac{M^{(0)}}{M^{(q)}} \sum_{m=1}^{M^{(q)}} \left[A_{m,*}^{(q)} \left(1 - \frac{y_m^{(q)}}{f_m^{(q)}(x^{(q)})} \right) \right] + \nabla S^{(q)}(x^{(q)}) \qquad (3.33)$$

$$g^{(q+1)} \leftarrow \frac{M^{(0)}}{M^{(q+1)}} \sum_{m=1}^{M^{(q+1)}} \left[A_{m,*}^{(q+1)} \left(1 - \frac{y_m^{(q+1)}}{f_m^{(q+1)}(x^{(q+1)})} \right) \right] + \nabla S^{(q+1)}(x^{(q+1)}) \ , \qquad (3.34)$$

where $A_{m,*}$ denotes the m^{th} row of the matrix A. With this choice of coarse scale cost functions, multigrid inversion works by the procedure specified in Fig. 3.1.

3.3 Adaptive Computation Allocation

The MultigridV subroutine in Fig. 3.1(b) specifies that $\nu_1^{(q)}$ fixed grid iterations are performed before each coarse grid update, and $\nu_2^{(q)}$ iterations are performed after the update. The convergence speed of the algorithm can be tuned through the choices of $\nu_1^{(q)}$ and $\nu_2^{(q)}$ at each scale. In practice, the best choice of these parameters also varies with the number of MultigridV iterations. For example, coarse fixed-grid optimization is typically more important in initial iterations, while fine fixed-grid optimization is more important during later iterations when the solution is close to its final value. For this reason, we can further improve convergence speed by adaptively changing the values of $\nu_1^{(q)}$ and $\nu_2^{(q)}$ with time instead of fixing the parameters to pre-determined values.

In this section, we describe how to adaptively allocate computation to the scale at which the algorithm can best reduce the cost [61]. In our adaptive scheme, we do not fix the $\nu_1^{(q)}$ and $\nu_2^{(q)}$ parameters in advance. Instead we perform fixed-grid updates as long as they continue to effectively reduce cost. This adaptive approach can further improve convergence speed and eliminates the need to select these parameters.

First, we would like the image updates to begin at the coarsest scale since this is usually more effective when the solution is far from the optimum. To do this, we initially set $\nu_1^{(q)} = 0$, so that when proceeding from fine to coarse scale in the first multigrid-V cycle we do not update the image and only update the r vector.

Second, when proceeding from coarse to fine scale in the first multigrid-V cycle, we perform the fixed-grid iterations until the change in the cost function falls below a threshold. More specifically, fixed-grid iterations are applied as long as the condition

$$C_1 : \quad \Delta c^{(q)} \geq \Delta_{max} c^{(q)} \, T \tag{3.35}$$

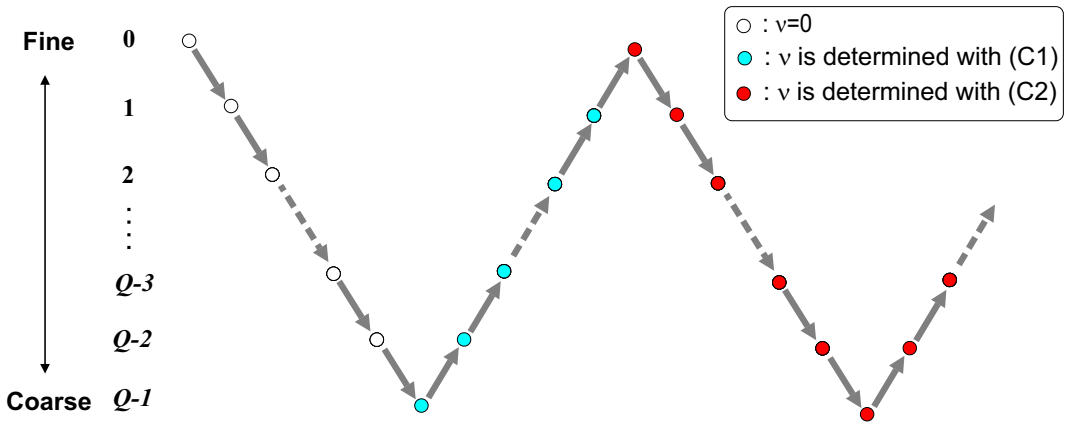


Fig. 3.2. Adaptive multigrid-V scheme

is satisfied, where $\Delta c^{(q)}$ is a state variable containing the reduction in cost that resulted from the most recent application of the fixed grid optimization at grid resolution q, $\Delta_{max}c^{(q)}$ is a state variable containing the maximum value that $\Delta c^{(q)}$ has taken on, and T is a threshold which we set to the value 0.1 in this paper. If the condition is not satisfied, the algorithm proceeds to the next scale.

Once the first multigrid cycle is complete, the adaptive multigrid algorithm compares the computational efficiency at the current scale q and at the next grid scale denoted by q_{next}, and performs the fixed grid iteration at scale q only if it is likely to be more effective than moving to scale q_{next}. More specifically, before each fixed-grid update, a conditional test, C_2, is evaluated. If the test is true, the fixed-grid update is performed; but if it is false, then the algorithm preceeds to the next grid scale q_{next}. This condition is given by

$$C_2: \quad \frac{\Delta c^{(q)}}{comp^{(q)}} \geq \frac{\Delta c^{(q_{next})}}{comp^{(q_{next})}}, \tag{3.36}$$

where $comp^{(q)}$ is the computation required for a single fixed-grid update at scale q. Importantly, since $\Delta c^{(q)}$ and $\Delta c^{(q_{next})}$ are state variables, these values are saved from the previous pass through grid resolutions q and q_{next}.

The adaptive MultigridV algorithm is schematically summarized in Fig. 3.2. While some adaptive multigrid algorithms have been developed for PDE solvers [106],

our adaptive scheme is unique because it uses the cost change as the criterion for adaptation. This is possible because our multigrid inversion method is based on an optimization framework [56,91], in contrast to conventional multigrid methods which are formulated as equation solvers.

3.4 Applications to Bayesian Emission and Transmission Tomography

In this section we apply the proposed multigrid inversion method to iterative reconstruction for emission and transmission tomography. The algorithms are formulated in a Bayesian reconstruction framework using both the quadratic data term and the Poisson noise model.

3.4.1 Multigrid tomographic inversion with quadratic data term

Emission tomography and transmission tomography use projected photon counts y to reconstruct the image x, which consists of a cross-sectional emission rate map and a cross-sectional attenuation map, respectively. The MAP image reconstruction problem is reduced to a minimization problem with the cost function [10, 101]

$$||\gamma - Px||_{\Lambda}^2 + S(x) \ , \tag{3.37}$$

where for the emission case we have

$$\gamma_m = y_m \tag{3.38}$$

$$\Lambda = \frac{1}{2} \ \mathrm{diag} \left\{ \frac{1}{y_1}, \frac{1}{y_2}, \ldots, \frac{1}{y_M} \right\} \ , \tag{3.39}$$

and for the transmission case we have

$$\gamma_m = \log \frac{y_T}{y_m} \tag{3.40}$$

$$\Lambda = \frac{1}{2} \ \mathrm{diag}\{y_1, y_2, \ldots, y_M\} \ , \tag{3.41}$$

where P is the forward projection matrix, y_T is the photon dosage per ray in the transmission case, and γ plays a role similar to y in (3.3).

Notice that since (3.37) has the form of (3.3), we can use the multigrid inversion algorithm decribed in Section 3.2.1 to compute the MAP reconstruction. However, to do this we must specify the coarse scale forward models, $f^{(q)}(\cdot)$, and the coarse scale stabalizing functions, $S^{(q)}(\cdot)$.

The fine scale forward model is given by the linear transformation

$$f(x) = Px \ . \tag{3.42}$$

The coarse scale forward model also has the linear form

$$f^{(q)}(x^{(q)}) = P^{(q)}x^{(q)} \ , \tag{3.43}$$

where $P^{(q)}$ is an $M^{(q)} \times N^{(q)}$ coarse scale projection matrix given by

$$P^{(q+1)} \triangleq J_{(q)}^{(q+1)} P^{(q)} I_{(q+1)}^{(q)} \ . \tag{3.44}$$

Note that $P^{(q+1)}$ in (3.44) can be pre-computed and stored since it is independent of the images.

Although in principle our multigrid inversion framework can work with any choice of data domain interpolator $J_{(q+1)}^{(q)}$ and decimator $J_{(q)}^{(q+1)}$, we need to choose them carefully to retain computational efficiency. We choose $J_{(q+1)}^{(q)}$ so that each row has only one non-zero element, and thus the resulting coarse scale weight matrix $\Lambda^{(q)}$ given by (3.22) is diagonal. For this reason, we interpolate using pixel replication along both the displacement and angle dimensions of the sinogram data. In other words, $J_{(q+1)}^{(q)}$ interpolates the sinogram data with the 1-D interpolation matrix

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \tag{3.45}$$

along both the angle and displacement axes. We choose the decimator to have the adjoint form of the interpolator, giving

$$J_{(q)}^{(q+1)} = \frac{1}{2}\left[J_{(q+1)}^{(q)}\right]^T \ . \tag{3.46}$$

Note that some other interpolation matrices, including the popular bilinear interpolator, do not preserve the sparsity of weight matrix $\Lambda^{(q)}$ at coarse scales.

For the image prior model we use the generalized Gaussian Markov random field (GGMRF) model [76], which is known to effectively enforce smoothness while preserving edges in tomographic reconstruction. In this case, the stabilizing function is given by

$$S(x) = \frac{1}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left|x_i - x_j\right|^p \ , \tag{3.47}$$

where σ is a normalization parameter, $1 \leq p \leq 2$ controls the degree of edge smoothness, the set $\mathcal{N}$ consists of all pairs of adjacent pixels, and b_{i-j} is a weight given to the pair of pixels i and j. We use the corresponding coarse scale stabilizing functions [91]

$$S^{(q)}(x^{(q)}) = \frac{1}{p(\sigma^{(q)})^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left|x_i^{(q)} - x_j^{(q)}\right|^p \ , \tag{3.48}$$

where $\sigma^{(q)}$ is given by $\sigma^{(q)} = 2^{q(1-\frac{d}{p})} \cdot \sigma^{(0)}$, and d is the dimensionality of the problem. The gradient terms of the stabilizing function used in (3.17), (3.18), (3.33), and (3.34) are computed by

$$\frac{\partial}{\partial x_n} S(x) = \frac{1}{(\sigma^{(q)})^p} \sum_{j\in\mathcal{N}_n} b_{n-j} |x_n^{(q)} - x_j^{(q)}|^{p-1} \mathrm{sgn}(x_n^{(q)} - x_j^{(q)}) \ . \tag{3.49}$$

3.4.2 Multigrid tomographic inversion for Poisson data model

In the emission case, the photon count Y_m for the m^{th} detector or detector pair is known to be described by the Poisson distribution (3.24) with mean and variance

$$f_m(x) = P_{m,*}x \ , \tag{3.50}$$

where $P_{m,*}$ is the m^{th} row of the matrix P. For this case, the MAP image reconstruction problem is reduced to minimizing the cost function (3.25) with the Poisson mean (3.50). We also use the coarse scale projection matrix of (3.44).

A similar method can be used for the transmission case, but with the Poisson mean given by

$$f_m(x) = y_T \exp(-P_{m,*}x) \ . \tag{3.51}$$

We use the coarse scale Poisson mean vector computed by

$$f_m^{(q)}(x^{(q)}) \quad = \quad y_T \exp(-P_{m,*}^{(q)} x^{(q)}) \tag{3.52}$$

where $P^{(q)}$ is once again given by (3.44).

Both emission and transmission cases use the same interpolation/decimation matrices and coarse scale stabilizing functions as described in Sec. 3.4.1.

3.5 Numerical Results

In this section, we compare three algorithms: the proposed multigrid algorithms with variable data resolution; the multigrid algorithms with fixed data resolution; and the fixed-grid ICD algorithm [10, 101]. We tested the algorithms for Bayesian reconstruction in emission and transmission tomography using the modified Shepp-Logan phantom [107] shown in Fig. 3.3(a). The width and the height of the bounding rectangle was 20 cm, and the two-dimensional region was discretized with 513×513 pixels. In the emission case, the brighter regions correspond to higher emission; and in the transmission case, the brighter regions correspond to higher absorption, with a peak absorption coefficient of 0.05 cm^{-1}. Projection data was simulated using 180 uniformly spaced angles, each with 512 uniformly spaced projections. The projection beam was assumed to have a triangular beam profile with a width of two times the projection spacing. In the emission case the total photon count per projection data was approximately 1.68×10^6 photons. In the transmission case, the dosage y_T per ray was 800 photons. Measurements were simulated as independent Poisson random

variables. The same data set was used for both the quadratic data term-based reconstruction and the Poisson model-based reconstruction.

Reconstructions were performed on 513×513 pixels. All three algorithms were initialized with the convolution backprojection (CBP) reconstructions shown in Fig. 3.3 (b) and (c). The CBP algorithm was implemented for a generalized Hamming reconstruction filter with frequency response $H(\omega) = H_{id}(\omega)(0.5 + 0.5\cos(\pi\omega/\omega_c))$ for $|\omega| < \omega_c$, where $H_{id}(\omega)$ is the ideal ramp filter. The cutoff frequency ω_c was chosen to yield minimum image root-mean-square error (RMSE), which was $\omega_c = 0.6\pi$ for transmission tomography and $\omega_c = 0.5\pi$ for emission tomography.

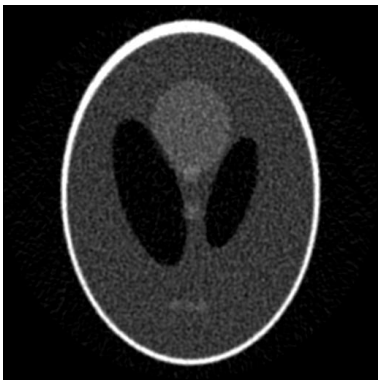
Both multigrid algorithms used a three level multigrid-V recursion, and used the fixed-grid ICD algorithm [10, 101] with random-order pixel updates. We chose the ν parameters in Fig. 3.1(b), which control the number of fixed-grid update iterations at each scale, adaptively, as described in Sec. 3.3. For fair comparison, we scaled the iteration number by the theoretical computational complexity. A detailed description for the conversion can be found in the Appendix C. The CBP computation is not included in the computational complexity since the CBP initialization is of negligible cost compared with the ensuing computation.

The image prior model parameters used were an eight point neighborhood GGMRF with $p = 1.2$, and $b_{j-k} = (2 - \sqrt{2})/4$ for nearest neighbors and $b_{j-k} = (\sqrt{2} - 1)/4$ for diagonal neighbors. We chose the image prior variance parameter to be $\sigma = 0.0025$ cm^{-1} in the transmission case and $\sigma = 0.4$ cm^{-1} in the emission case. These values were lower than the optimal parameters yielding minimum image RMSE, but they resulted in qualitatively better reconstructions in spite of a slightly larger RMSE.

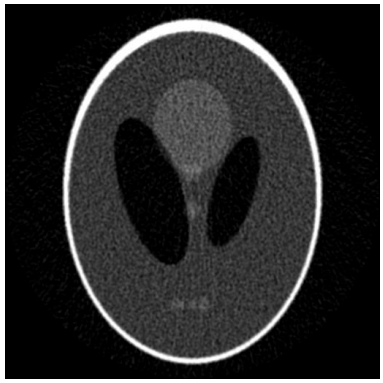
Figures 3.4(a), 3.5(a), 3.6(a), and 3.7(a) compare the convergence speed of the algorithms in terms of the cost function. For both imaging modalities and both data likelihood functions, the multigrid algorithm with variable data resolution converged twice as fast as the multigrid algorithm with fixed data resolution. Importantly, although the convergence of the fixed grid ICD algorithms in the initial few iterations



(a)



(b)



(c)

Fig. 3.3. (a) true phantom (b) CBP reconstruction for emission tomography (c) CBP reconstruction for transmission tomography

is comparable with that of the multigrid algorithms with fixed data resolution, they eventually require many more iterations ($30 \sim 50$ iterations) to reduce the cost to the value to which the multigrid algorithms with variable data resolution converged in $5 \sim 8$ iterations.

Figures 3.4(b), 3.5(b), 3.6(b), and 3.7(b) compare the convergence speed of the algorithms in terms of RMSE of reconstructed images. For all the cases, the multigrid algorithm with variable data resolution converged fastest. The fixed-grid algorithm behaved poorly at the first iteration, and it produced some salt and pepper noise by overshooting in some image pixel updates. Again, the fixed-grid algorithm required about $30 \sim 50$ iterations to reduce image RMSE to the value that the multigrid algorithms converged to in $5 \sim 8$ iterations. Since the convexity of the cost function excludes the possibility of being trapped into a local minimum, the difference in convergence speed is probably due to the fact that there are some error components which the fixed-grid optimization cannot effectively remove.

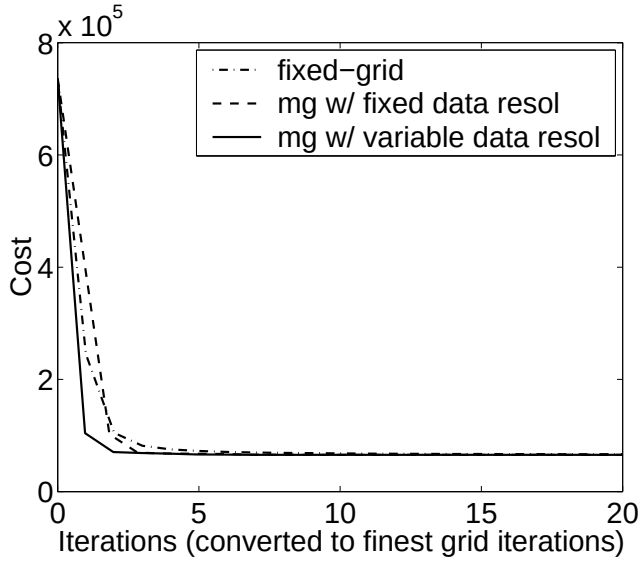
The convergence plots show that all the algorithms eventually converged to the same cost and RMSE, which should be a natural consequence of the convex optimization function. However, although the cost decrease rate of the multigrid algorithms and the fixed-grid algorithm are similar for the initial iterations, the RMSE convergence results indicate that they converged following different optimization trajectories. The trajectory of the multigrid algorithms are perhaps more favorable because they yielded significantly smaller RMSE image error before full convergence.

Figures 3.8 and 3.9 show the reconstructed images for emission tomography with the Poisson noise model and the quadratic approximation of data likelihood, respectively, and Figs 3.10 and 3.11 show the reconstructed images for transmission tomography. For all cases, the final reconstruction quality was quantitatively and qualitatively almost the same for the three algorithms. However, the fixed-grid algorithm yielded poorer image quality even with twice or four times the computation that the multigrid methods required to converge. For example, the fixed-grid reconstruction in Fig. 3.9(b) and (c) with 14 and 28 iterations, respectively, was visually

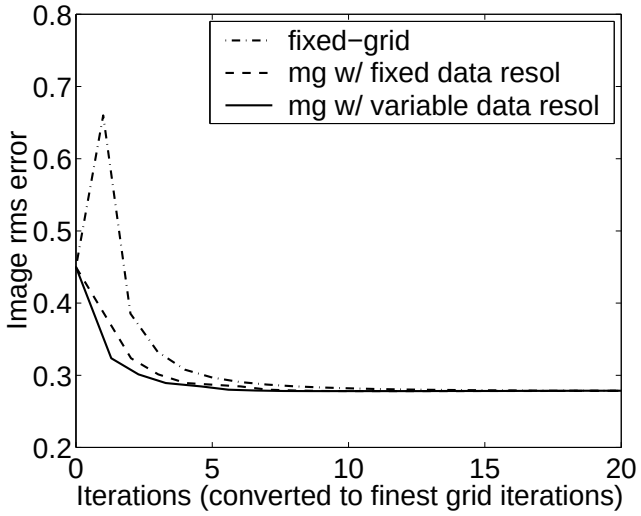
worse than the multigrid reconstructions with only 5.31 or 8.06 iterations, which are shown in Fig. 3.9(e) and (f). The reconstructions by all the statistical methods improve the image quality compared to the CBP reconstruction. In summary, the proposed multigrid algorithm significantly saved computations compared with the fixed-grid ICD algorithm initialized with the CBP reconstruction.

3.6 Conclusions

The multigrid inversion methods with variable resolution data and image spaces were proposed. In formulating a set of optimization functions at different scales, the algorithm changes grid resolution of both measurement data space and image space, and thus improves computational efficiency further than the previous multigrid inversion methods which changes resolutions in the image space only. Application to conventional transmission and emission tomography problems demonstrated substantially reduced computation relative to the fixed-grid ICD algorithm and our previous multigrid inversion with fixed data resolution.

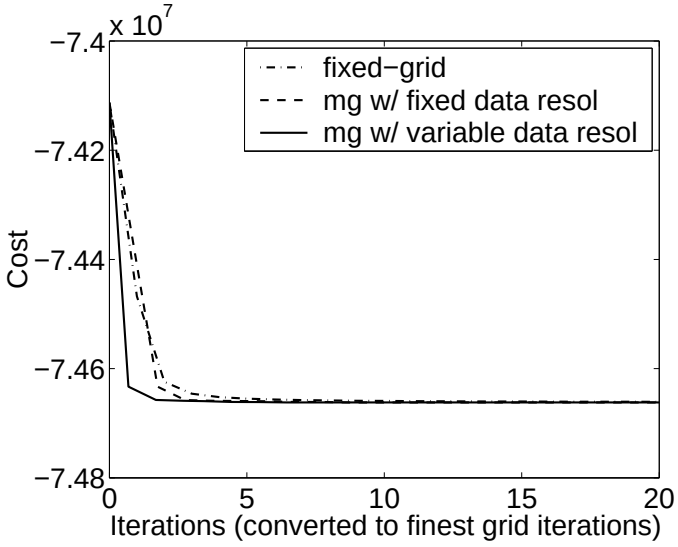


(a)

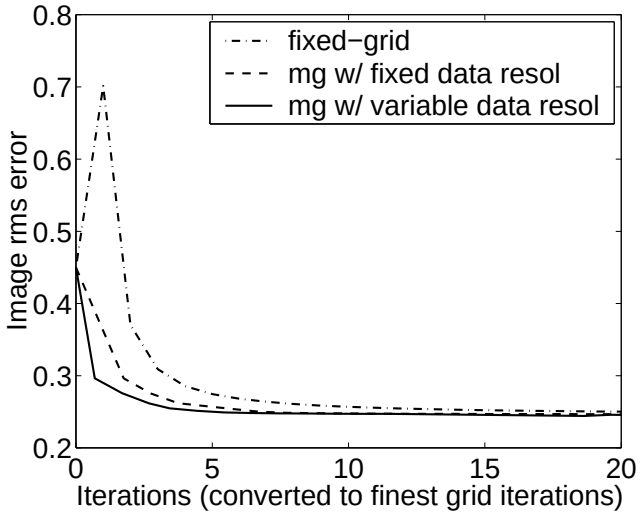


(b)

Fig. 3.4. Convergence in emission tomography with quadratic data term in terms of (a) cost function and (b) image rms error

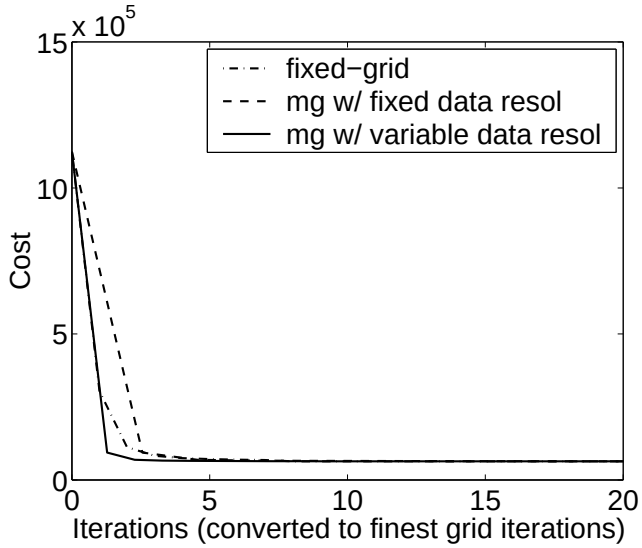


(a)

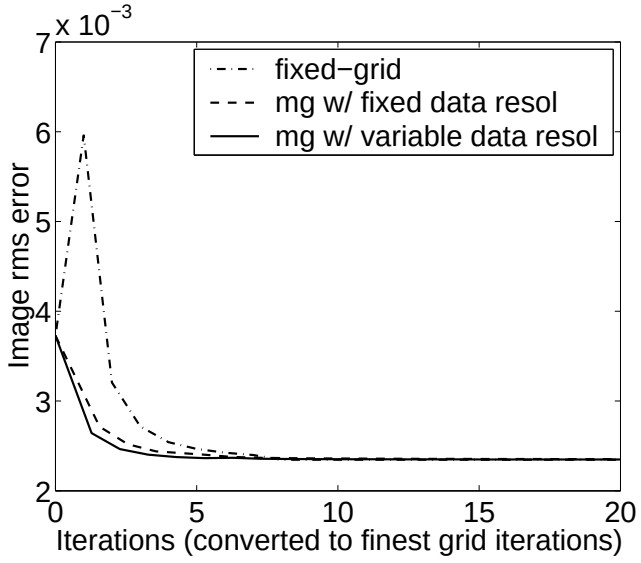


(b)

Fig. 3.5. Convergence in emission tomography with the Poisson noise model in terms of (a) cost function and (b) image rms error

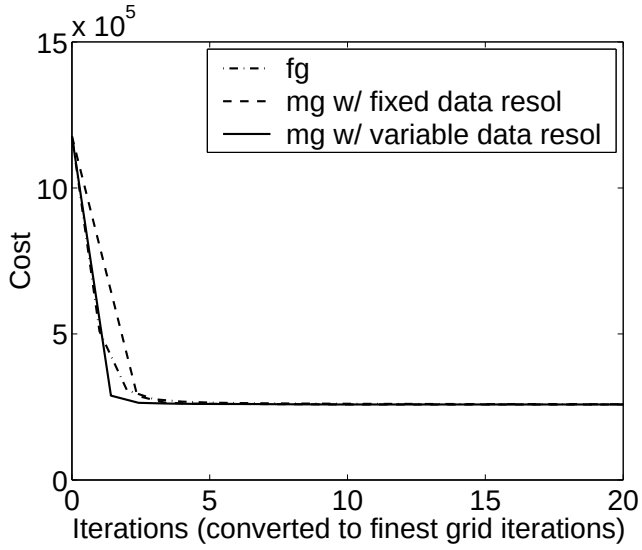


(a)

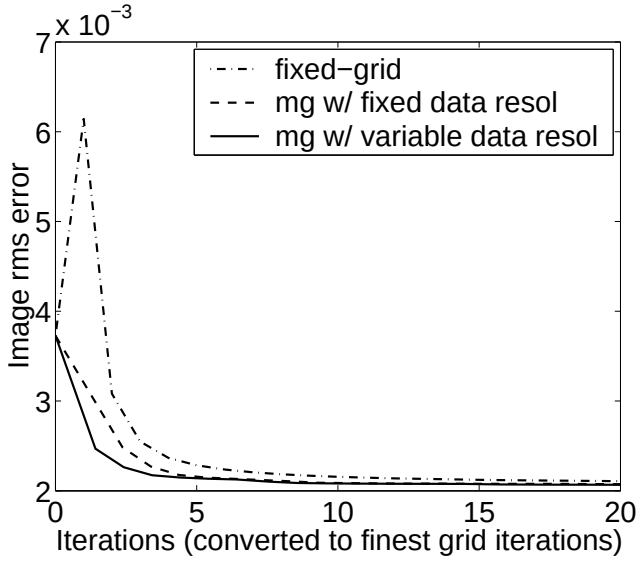


(b)

Fig. 3.6. Convergence in transmission tomography with quadratic data term in terms of (a) cost function and (b) image rms error

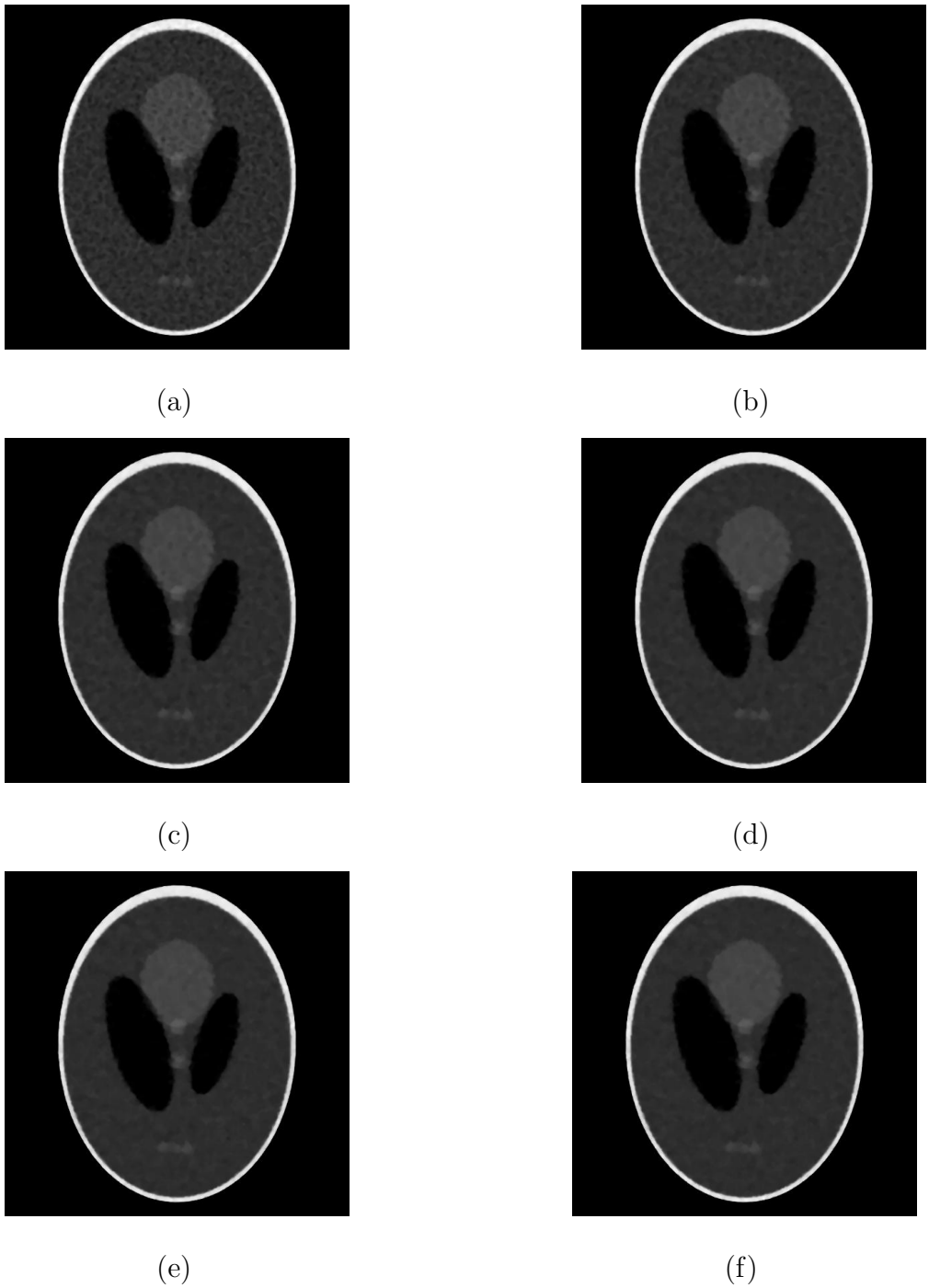


(a)



(b)

Fig. 3.7. Convergence in transmission tomography with the Poisson noise model in terms of (a) cost function and (b) image rms error

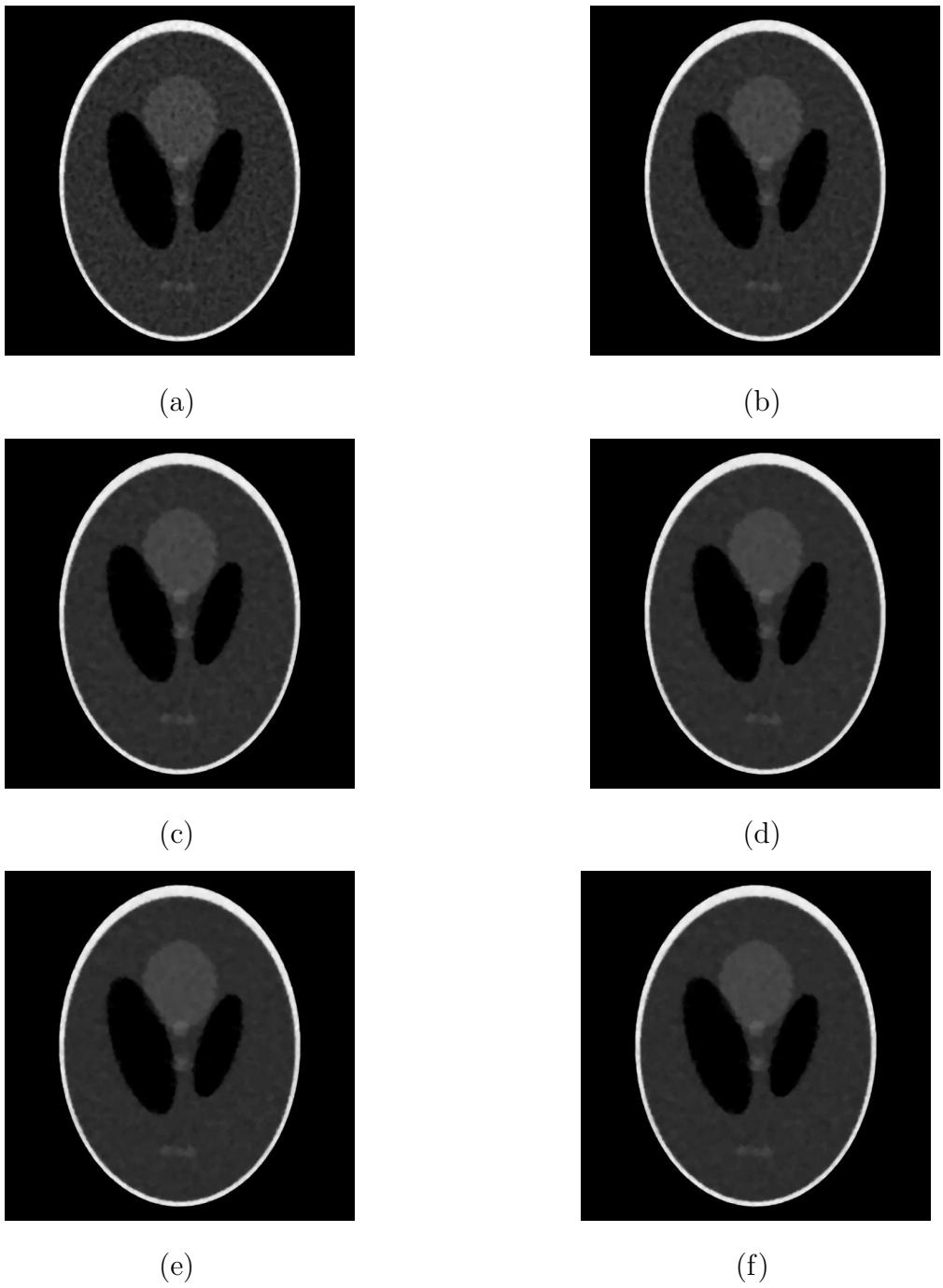


(a) (b)

(c) (d)

(e) (f)

Fig. 3.8. Reconstructions for emission tomography with quadratic data term: fixed-grid algorithm with (a) 7 iterations (b) 14 iterations (c) 28 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (7.79 iterations); and (f) multigrid algorithm with variable data resolution (5.94 iterations)

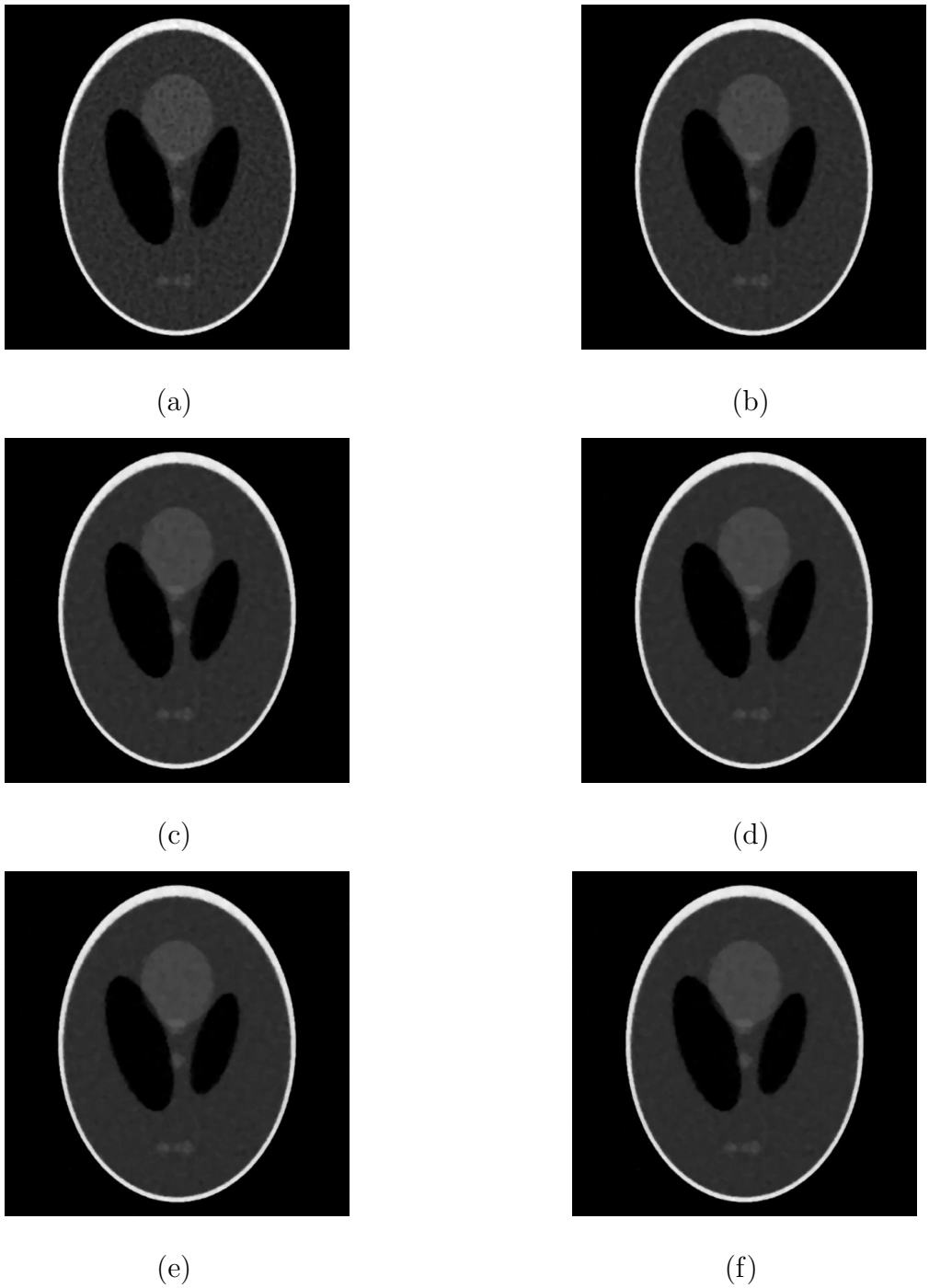


(a) (b)

(c) (d)

(e) (f)

Fig. 3.9. Reconstructions for emission tomography with the Poisson noise model: fixed-grid algorithm with (a) 7 iterations (b) 14 iterations (c) 28 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (8.06 iterations); and (f) multigrid algorithm with variable data resolution (5.31 iterations)



(a) (b)

(c) (d)

(e) (f)

Fig. 3.10. Reconstructions for transmission tomography with quadratic data term: fixed-grid algorithm with (a) 7 iterations (b) 14 iterations (c) 28 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (7.48 iterations); and (f) multigrid algorithm with variable data resolution (5.81 iterations)



(a) (b)



(c) (d)



(e) (f)

Fig. 3.11. Reconstructions for transmission tomography with the Poisson noise model: fixed-grid algorithm with (a) 8 iterations (b) 16 iterations (c) 32 iterations and (d) 50 iterations; (e) multigrid algorithm with fixed data resolution (9.06 iterations); and (f) multigrid algorithm with variable data resolution (6.46 iterations)

4. SOURCE-DETECTOR CALIBRATION IN THREE-DIMENSIONAL BAYESIAN OPTICAL DIFFUSION TOMOGRAPHY

4.1 Introduction

Optical diffusion tomography (ODT) is an imaging modality that has potential in applications such as medical imaging, environmental sensing, and non-destructive testing [2]. In this technique, measurements of the light that propagates through a highly scattering medium are used to reconstruct the absorption and/or the scattering properties of the medium as a function of position. In highly scattering media such as tissue, the diffusion approximation to the transport equations is sufficiently accurate and provides a computationally tractable forward model. However, the inverse problem of reconstructing the absorption and/or scattering coefficients from measurements of the scattered light is highly nonlinear. This nonlinear inverse problem can be very computationally expensive, so methods that reduce the computational burden are of critical importance [56, 63, 64, 77, 108].

Another important issue for practical ODT imaging, that is addressed in this paper, is accurate modeling of the source and detector coupling coefficients [109]. These coupling coefficients determine weights for sources and detectors in a diffusion equation model for the scattering domain. The physical source of the source/detector coupling variability is associated with the optical components external to the scattering domain, for example, the placement of fibers, the variability in switches, etc. Variations in the coupling coefficients can result in severe, systematic reconstruc-

tion distortions. In spite of its practical importance, this issue has received little attention.

Two preprocessing methods have been investigated to correct for source/detector coupling errors before inversion. Jiang *et al.* [110,111] calibrated coupling coefficients and a boundary coefficient by comparing prior measurements of photon flux density for a homogeneous medium with the corresponding computed values. This scheme has been applied in clinical studies [112–114]. This method of calibration requires a set of reference measurements from a homogeneous sample, in addition to the measurements used to reconstruct the inhomogeneous image. Iftimia *et al.* [115] proposed a preprocessing scheme that involved minimization of the mean square error between the measurements for the given inhomogeneous phantom and the computed values with an assumed homogeneous medium. However, although this approach does not require prior homogeneous reference measurements, it neglects the influence of an inhomogeneous domain when determining the source and detector weights.

In order to reconstruct the image from a single set of measurements from the domain to be imaged, it is necessary to estimate the coupling coefficients as the image is reconstructed. For example, Boas *et al.* [109] proposed a scheme for estimating individual coupling coefficients as part of the reconstruction process. They simultaneously estimated both absorption and coupling coefficients by formulating a linear system which consisted of the perturbations of the measurements in a Rytov approximation and the logarithms of the source and detector coupling coefficients. No results have been reported for nonlinear reconstruction of both absorption and diffusion images, and the individual coupling coefficients.

In this paper, we describe an efficient algorithm for estimating individual source and detector coupling coefficients as part of the reconstruction process for both absorption and diffusion images. This approach is based on the formulation of our problem in a unified Bayesian regularization framework containing terms for both the unknown 3-D optical properties and the coupling coefficients. The resulting cost function is then jointly minimized to both reconstruct the image and estimate the

needed coefficients. To perform this minimization, we adapt our iterative coordinate decent optimization method [77] to include closed form steps for the update of the coupling coefficient estimates. This unified optimization approach results in an algorithm which can reconstruct images and estimate the coupling coefficients without the need for prior calibration. In a previous experiment, we used the algorithm to effectively estimate a single coefficient from a measured 3-D data set [13]. Simulation results show that our method can substantially improve reconstruction quality even when there are a large number of severely non-uniform coupling coefficients. Our approach is applied to a simple phantom experiment.

4.2 Problem Formulation

In a highly scattering medium with low absorption, such as soft tissue in the 650-1300 nm wavelength range, the photon flux density is accurately modeled by the diffusion equation [116,117]. In frequency domain optical diffusion imaging, the light source is amplitude modulated at angular frequency ω, and the complex modulation envelope of the optical flux density is measured at the detectors. The complex amplitude $\phi_k(r)$ of the modulation envelope due to a point source at position a_k satisfies the frequency domain diffusion equation

$$\nabla \cdot [D(r)\nabla\phi_k(r)] + [-\mu_a(r) - j\omega/c]\,\phi_k(r) = -\delta(r - a_k), \tag{4.1}$$

where r is position, c is the speed of light in the medium, $D(r)$ is the diffusion coefficient, and $\mu_a(r)$ is the absorption coefficient. We consider a region to be imaged that is surrounded by K point sources at positions a_k, for $1 \leq k \leq K$, and M detectors at positions b_m, for $1 \leq m \leq M$. The 3-D domain is discretized into N grid points, denoted by $r_1, \cdots, r_N$. The unknown image is then represented by a $2N$ dimensional column vector x containing the absorption and diffusion coefficients at each discrete grid point

$$x = [\mu_a(r_1), \ldots, \mu_a(r_N), D(r_1), \ldots, D(r_N)]^t \ . \tag{4.2}$$

We will use the notation $\phi_k(r;x)$ in place of $\phi_k(r)$, in order to emphasize the dependence of the solution to (4.1) on the unknown material properties x.

Let y_{km} be the complex measurement at detector location b_m and using a source at location a_k. This measurement is a sample of a random variable Y_{km}, which we will model as a sum of the true signal and Gaussian noise. The datum mean value of Y_{km} is given by

$$E[Y_{km}|x, s_k, d_m] = s_k d_m \phi_k(b_m; x) \ , \tag{4.3}$$

where $\phi_k(b_m;x)$ is the solution of (4.1) evaluated at position b_m; s_k and d_m are complex constants representing the unknown source and detector distortions; and $E[\cdot|x, s_k, d_m]$ denotes the conditional expectation given x, s_k, and d_m. [1]

Our objective is to simultaneously estimate the unknown image x together with the unknown source and detector coupling coefficient vectors $s = [s_1, s_2, \ldots, s_K]^t$ and $d = [d_1, d_2, \ldots, d_M]^t$. The coupling coefficients are different for different sources and detectors, and are not known *a priori*. In general, the values of s_k and d_m will vary in both amplitude and phase for real physical systems. Typically, amplitude variations can be caused by different excitation intensities for the sources and different collection efficiencies for the detectors, and phase variation can be caused by the different effective positions of the sources and detectors. Without these parameter vectors, accurate reconstruction of x is not possible.

The measurement vector y is formed by raster ordering the measurements y_{km} in the form

$$y = [y_{11}, \ldots, y_{1M}, y_{21}, \ldots, y_{2M}, \ldots, y_{KM}]^t \ . \tag{4.4}$$

The conditional expectation of $Y = [Y_{11}, \ldots, Y_{1M}, Y_{21}, \ldots, Y_{2M}, \ldots, Y_{KM}]^t$ is then given by

$$E[Y|x, s, d] = \mathrm{diag}(s \otimes d)\Phi(x) \ , \tag{4.5}$$

[1] We assume that the physical sources and detectors provide an adequate measure of ϕ, that they do not perturb the diffusion equation solution, and that they have an equivalent point representation.

where $s \otimes d$ is the Kronecker product of s and d, $\mathrm{diag}(w)$ is a diagonal matrix whose (i,i)-th element is equal to the i-th element of the vector w, and $\Phi(x)$ is the corresponding raster order of the values $\phi_k(b_m; x)$ given by

$$\Phi(x) = [\ \phi_1(b_1; x), \phi_1(b_2; x), \ldots, \phi_1(b_M; x), \phi_2(b_1; x), \ldots, \phi_K(b_M; x)\ ]^t\ . \tag{4.6}$$

In order to simplify notation, we define the forward model vector $f(x, s, d)$ as

$$f(x, s, d) = \mathrm{diag}(s \otimes d)\Phi(x)\ . \tag{4.7}$$

We use a shot noise model for the detector noise. [77,78] The shot noise model assumes independent noise measurements that are Gaussian with variance proportional to the signal amplitude. This results in the following expression for the conditional density of Y

$$p(y|x, s, d, \alpha) = \frac{1}{(\pi\alpha)^P |\Lambda|^{-1}} \exp\left[-\frac{||y - f(x, s, d)||^2_\Lambda}{\alpha}\right], \tag{4.8}$$

where $P = KM$ is the number of measurements, α is an unknown parameter that scales the noise variance, $\Lambda = \mathrm{diag}([1/|y_{11}|, \ldots, 1/|y_{1M}|, 1/|y_{21}|, \ldots, 1/|y_{KM}|]^t)$, and $||w||^2_\Lambda = w^H \Lambda w$.

We determine x, s, d, and α from the measurements y. Because this is an ill-posed inverse problem, we employ a Bayesian framework to incorporate a prior model for x, the image [77]. We then maximize the posterior probability of x jointly with respect to y, s, d, and α. This yields the estimators

$$\begin{aligned}(\hat{x}_{MAP}, \hat{s}, \hat{d}, \hat{\alpha}) &= \arg \max_{(x \geq 0, s, d, \alpha)} \{\ \log p(x|y, s, d, \alpha)\ \} \\ &= \arg \max_{(x \geq 0, s, d, \alpha)} \{\ \log p(y|x, s, d, \alpha) + \log p(x)\ \}\ ,\end{aligned} \tag{4.9}$$

where $p(y|x, s, d, \alpha)$ is the data likelihood, and $p(x)$ is the prior model for the image. The estimate $\hat{x}_{MAP}$ is essentially the maximum *a posteriori* (MAP) estimate of the image, but it is computed by simultaneously optimizing with respect to the unknown parameters s, d, and α. Quantities such as s, d, and α are sometimes known as nuisance parameters, because they are not of direct interest, but are required for

accurate estimation of the desired quantity x. A variety of methods have been proposed for estimating such parameters. These methods range from true maximum likelihood estimation using Monte Carlo Markov chain (MCMC) techniques [67, 118, 119], to joint MAP estimation of the unknown image and parameters [65, 66]. Our method is a form of joint MAP estimation, but with a uniform (i.e., improper) prior distribution for s, d, and α. It is worth noting that such estimators can behave poorly in certain cases [120]. However, when the number of measurements is large compared to the dimensionality of the unknowns, as in our case for s, d, and α, these estimators generally work well.

We use the generalized Gaussian Markov random field (GGMRF) prior model [76] for the image x,

$$\begin{aligned} p(x) &= p([\mu_a(r_1), \mu_a(r_2), \ldots, \mu_a(r_N)]^T) \cdot p([D(r_1), D(r_2), \ldots, D(r_N)]^T) \\ &= \left[\frac{1}{{\sigma_0}^N z(p_0)} \exp\left\{ -\frac{1}{p_0 {\sigma_0}^{p_0}} \sum_{\{i,j\}\in\mathcal{N}} b_{0,i-j} |x_i - x_j|^{p_0} \right\} \right] \\ &\quad \cdot \left[\frac{1}{{\sigma_1}^N z(p_1)} \exp\left\{ -\frac{1}{p_1 {\sigma_1}^{p_1}} \sum_{\{i,j\}\in\mathcal{N}} b_{1,i-j} |x_{N+i} - x_{N+j}|^{p_1} \right\} \right] \\ &= \prod_{u=0}^{1} \left[\frac{1}{{\sigma_u}^N z(p_u)} \exp\left\{ -\frac{1}{p_u {\sigma_u}^{p_u}} \sum_{\{i,j\}\in\mathcal{N}} b_{u,i-j} |x_{uN+i} - x_{uN+j}|^{p_u} \right\} \right] \end{aligned} \tag{4.10}$$

where σ_0 and σ_1 are normalization parameters for μ_a and D, respectively, and $1 \leq p_0 \leq 2$ and $1 \leq p_1 \leq 2$ control the degree of edge smoothness for μ_a and D, respectively. The set $\mathcal{N}$ consists of all pairs of adjacent grid points, $z(p_0)$ and $z(p_0)$ are normalization constants, and $b_{0,i-j}$ and $b_{1,i-j}$ represent the coefficients assigned to neighbors i and j for μ_a and D, respectively. This prior model enforces smoothness in the solution while preserving sharp edge transitions, and its effectiveness for this kind of problem has been shown previously [77].

4.3 Optimization

Let $c(x, s, d, \alpha)$ denote the cost function to be minimized in (4.9). Then using the models of (4.8) and (4.10) and removing constant terms results in

$$\begin{aligned} c(x, s, d, \alpha) &= \frac{1}{\alpha}||y - f(x, s, d)||_{\Lambda}^2 \\ &\quad + P \log \alpha + \sum_{u=0}^{1} \frac{1}{p_u {\sigma_u}^{p_u}} \sum_{\{i,j\} \in \mathcal{N}} b_{u,i-j} |x_{uN+i} - x_{uN+j}|^{p_u}. \end{aligned} \quad (4.11)$$

The objective is then to compute

$$(\hat{x}_{MAP}, \hat{s}, \hat{d}, \hat{\alpha}) = \arg \min_{(x \geq 0, s, d, \alpha)} c(x, s, d, \alpha) \ . \quad (4.12)$$

To solve this problem, we adapt the iterative coordinate decent (ICD) method [77]. The ICD method works by sequentially updating parameters of the optimization, so that each update monotonically reduces the cost function. Previous implementations of ICD sequentially updated pixels in the vector x. Here we generalize the ICD method so that the parameters s, d, and α are also included in the sequence of updates. More specifically, in each iteration of the ICD algorithm, s, d, α, and x are updated sequentially using the relations

$$\hat{\alpha} \leftarrow \arg\min_{\alpha} \ c(\hat{x}, \hat{s}, \hat{d}, \alpha) \quad (4.13)$$

$$\hat{s} \leftarrow \arg\min_{s} \ c(\hat{x}, s, \hat{d}, \hat{\alpha}) \quad (4.14)$$

$$\hat{d} \leftarrow \arg\min_{d} \ c(\hat{x}, \hat{s}, d, \hat{\alpha}) \quad (4.15)$$

$$\hat{x} \leftarrow ICD_update_x \left\{ c(x, \hat{s}, \hat{d}, \hat{\alpha}), \hat{x} \right\} \quad (4.16)$$

where the ICD_update_x operation performs one iteration of ICD optimization to reduce the cost function $c(\cdot, \hat{s}, \hat{d}, \hat{\alpha})$ starting at the initial value $\hat{x}$. The result of ICD_update_x is then used to update the value of $\hat{x}$. Iterative application of these update equations produces a convergent sequence of deceasing costs.

The updates of (4.13), (4.14), and (4.15) can be calculated in closed form by setting the partial derivative with respect to each variable to zero and solving the resulting equations to yield

$$\hat{\alpha} \leftarrow \frac{1}{P} \ || \ y - f(\hat{x}, \hat{s}, \hat{d}) \ ||_{\Lambda}^2 \quad (4.17)$$

$$\hat{s}_k \leftarrow \frac{[\ \mathrm{diag}(\hat{d})\ \Phi_k^{(s)}(\hat{x})\ ]^H \Lambda_k^{(s)}\ y}{||\ \mathrm{diag}(\hat{d})\ \Phi_k^{(s)}(\hat{x})\ ||^2_{\Lambda_k^{(s)}}} \qquad k = 1, 2, \ldots, K \tag{4.18}$$

$$\hat{d}_m \leftarrow \frac{[\ \mathrm{diag}(\hat{s})\ \Phi_m^{(d)}(\hat{x})\ ]^H \Lambda_m^{(d)}\ y}{||\ \mathrm{diag}(\hat{s})\ \Phi_m^{(d)}(\hat{x})\ ||^2_{\Lambda_m^{(d)}}} \qquad m = 1, 2, \ldots, M, \tag{4.19}$$

where $\Lambda_k^{(s)} = \mathrm{diag}(\ [\ 1/|y_{k1}|, 1/|y_{k2}|, \ldots, 1/|y_{kM}|\ ]^t\ )$ and $\Lambda_m^{(d)} = \mathrm{diag}(\ [\ 1/|y_{1m}|, 1/|y_{2m}|, \ldots, 1/|y_{Km}|\ ]^t\ )$ are the inverse diagonal covariance matrices associated with source k and detector m, respectively, $\Phi_k^{(s)}(\hat{x}) = [\ \phi_k(b_1; \hat{x}), \phi_k(b_2; \hat{x}), \ldots, \phi_k(b_M; \hat{x})\ ]^t$ and $\Phi_m^{(d)}(\hat{x}) = [\ \phi_1(b_m; \hat{x}), \phi_2(b_m; \hat{x}), \ldots, \phi_K(b_m; \hat{x})\ ]^t$ are the complex amplitude vectors associated with source k and detector m, respectively, and H denotes the Hermitian transpose.

The update of the variable x in (4.16) is of course more difficult since x is a high dimensional vector, particularly in the 3-D case. To update the image, we use one scan of the ICD algorithm as an ICD_update_x operation. One ICD scan involves sequentially updating each element of x with random ordering, and incorporation of the updated elements as the scan progresses. During this scan each element of x is updated only once. At the beginning of an ICD scan, the nonlinear functional $f(x, s, d)$ is first expressed using a Taylor expansion as

$$||y - f(x, s, d)||^2_\Lambda \simeq ||y - f(\hat{x}, \hat{s}, \hat{d}) - f'(\hat{x}, \hat{s}, \hat{d})\Delta x||^2_\Lambda\ , \tag{4.20}$$

where $\Delta x = x - \hat{x}$, and $f'(\hat{x}, \hat{s}, \hat{d})$ represents the Fréchet derivative of $f(x, \hat{s}, \hat{d})$ with respect to x at $x = \hat{x}$. Using (4.20), an approximate cost function for the original problem is

$$c(x, \hat{s}, \hat{d}, \hat{\alpha}) \simeq \frac{1}{\hat{\alpha}}||z - f'(\hat{x}, \hat{s}, \hat{d})x||^2_\Lambda + \sum_{u=0}^{1} \frac{1}{p_u \sigma_u{}^{p_u}} \sum_{\{i,j\} \in \mathcal{N}} b_{u,i-j} |x_{uN+i} - x_{uN+j}|^{p_u} \tag{4.21}$$

where

$$z \ = \ y - f(\hat{x}, \hat{s}, \hat{d}) + f'(\hat{x}, \hat{s}, \hat{d})\hat{x}\ . \tag{4.22}$$

Then, with the other image elements fixed, the ICD update for $\hat{x}_{uN+i}$ is given by

$$\hat{x}_{uN+i} \leftarrow \arg \min_{x_{uN+i} \geq 0} \left\{ \frac{1}{\hat{\alpha}} \left|\left| y - f(\hat{x}, \hat{s}, \hat{d}) - \left[f'(\hat{x}, \hat{s}, \hat{d}) \right]_{*(uN+i)} (x_{uN+i} - \hat{x}_{uN+i}) \right|\right|^2_\Lambda \right.$$

$$+\frac{1}{p_u \sigma^{p_u}} \sum_{j \in \mathcal{N}_i} b_{u,i-j} |x_{uN+i} - \hat{x}_{uN+j}|^{p_u} \Bigg\} \ , \tag{4.23}$$

where $[f'(\hat{x}, \hat{s}, \hat{d})]_{*(uN+i)}$ is the $(uN+i)$-th column of Fréchet matrix, and $\mathcal{N}_i$ is the set of grid points neighboring grid point i. To compute the solution to (4.23), we express the first term as a quadratic function of x_{uN+i} and then perform a one-dimensional minimization that is solved by a half-interval search for the root of the analytical derivative [77].

The Fréchet derivative $f'(\hat{x}, \hat{s}, \hat{d})$ is a $P \times 2N$ complex matrix given by

$$f'(\hat{x}, \hat{s}, \hat{d})$$

$$= \begin{bmatrix} \frac{\partial f_{11}(\hat{x},\hat{s}_1,\hat{d}_1)}{\partial \mu_a(r_1)} & \cdots & \frac{\partial f_{11}(\hat{x},\hat{s}_1,\hat{d}_1)}{\partial \mu_a(r_N)} & \frac{\partial f_{11}(\hat{x},\hat{s}_1,\hat{d}_1)}{\partial D(r_1)} & \cdots & \frac{\partial f_{11}(\hat{x},\hat{s}_1,\hat{d}_1)}{\partial D(r_N)} \\ \frac{\partial f_{12}(\hat{x},\hat{s}_1,\hat{d}_2)}{\partial \mu_a(r_1)} & \cdots & \frac{\partial f_{12}(\hat{x},\hat{s}_1,\hat{d}_2)}{\partial \mu_a(r_N)} & \frac{\partial f_{12}(\hat{x},\hat{s}_1,\hat{d}_2)}{\partial D(r_1)} & \cdots & \frac{\partial f_{12}(\hat{x},\hat{s}_1,\hat{d}_2)}{\partial D(r_N)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{1M}(\hat{x},\hat{s}_1,\hat{d}_M)}{\partial \mu_a(r_1)} & \cdots & \frac{\partial f_{1M}(\hat{x},\hat{s}_1,\hat{d}_M)}{\partial \mu_a(r_N)} & \frac{\partial f_{1M}(\hat{x},\hat{s}_1,\hat{d}_M)}{\partial D(r_1)} & \cdots & \frac{\partial f_{1M}(\hat{x},\hat{s}_1,\hat{d}_M)}{\partial D(r_N)} \\ \frac{\partial f_{21}(\hat{x},\hat{s}_2,\hat{d}_1)}{\partial \mu_a(r_1)} & \cdots & \frac{\partial f_{21}(\hat{x},\hat{s}_2,\hat{d}_1)}{\partial \mu_a(r_N)} & \frac{\partial f_{21}(\hat{x},\hat{s}_2,\hat{d}_1)}{\partial D(r_1)} & \cdots & \frac{\partial f_{21}(\hat{x},\hat{s}_2,\hat{d}_1)}{\partial D(r_N)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{KM}(\hat{x},\hat{s}_K,\hat{d}_M)}{\partial \mu_a(r_1)} & \cdots & \frac{\partial f_{KM}(\hat{x},\hat{s}_K,\hat{d}_M)}{\partial \mu_a(r_N)} & \frac{\partial f_{KM}(\hat{x},\hat{s}_K,\hat{d}_M)}{\partial D(r_1)} & \cdots & \frac{\partial f_{KM}(\hat{x},\hat{s}_K,\hat{d}_M)}{\partial D(r_N)} \end{bmatrix} , \tag{4.24}$$

where the first N columns correspond to the μ_a components of x and the remaining N columns correspond to the D components. In a similar manner to the Fréchet derivative commonly used for unity coupling coefficients [121], it can be shown that each element of the matrix is given by

$$\frac{\partial f_{km}(\hat{x}, \hat{s}_k, \hat{d}_m)}{\partial \mu_a(r_i)} = -\hat{s}_k \hat{d}_m g(b_m, r_i; \hat{x}) \phi_k(r_i; \hat{x}) A \tag{4.25}$$

$$\frac{\partial f_{km}(\hat{x}, \hat{s}_k, \hat{d}_m)}{\partial D(r_i)} = -\hat{s}_k \hat{d}_m \nabla g(b_m, r_i; \hat{x}) \cdot \nabla \phi_k(r_i; \hat{x}) A \ , \tag{4.26}$$

where A is the voxel volume, the Green's function $g(b_m, r_i; \hat{x})$ is the solution of (4.1) for a point source located at b_m (i.e., by setting $a_k \leftarrow b_m$ in (4.1), using reciprocity to reduce computation [121]) and a given image $\hat{x}$, ∇ is the gradient operator with

respect to r_i, and domain discretization errors are ignored. Note that the Fréchet derivative is the product of the coupling coefficient terms $\hat{s}_k \hat{d}_m$ and the derivative of $\phi_k(b_m; \hat{x})$ with respect to the optical parameter at that grid point. Thus, if the coupling coefficients are not accurately estimated, the formulas (4.25) and (4.26) do not yield accurate Fréchet derivatives, and thus the computed gradient direction of the cost function in (4.12) is not accurate. Therefore, accurate estimation of the coupling coefficients is essential for the ICD-Born iteration scheme.

The dimensions of the Fréchet derivative matrix are very large for practical 3-D imaging. For example, $(KM \times 2N \times 8) = 790$ MBytes of memory are needed to store the Fréchet derivative matrix for 30 sources, 48 detectors and a $33 \times 33 \times 33$ grid point image, if 4 bytes are used for a real number. However, the storage can be reduced by exploiting two facts. First, only the $(uN + i)$-th column of the Fréchet derivative matrix is needed to update x_{uN+i}, as seen in (4.23). Second, the Fréchet derivative in (4.25) and (4.26) is separable into the $\phi_k(r_i; \hat{x})$ term and the $g(b_m, r_i; \hat{x})$ term. Thus, we compute only $\phi_k(\,\cdot\,; \hat{x})$ for $k = 1, 2, \ldots, K$ and $g(b_m, \cdot\,; \hat{x})$ for $m = 1, 2, \ldots, M$ before the ICD update of the whole image, and then when x_i is updated the i-th column of the Fréchet derivative is computed using these vectors. This method, which involves storing the forward solutions for all sources, the Green's function for all detectors, and only one column of the Fréchet derivative matrix, reduces the required memory to $(KN + MN + KM) \times 8$ bytes without requiring additional computation. In the above example, the required memory is then only 22 MBytes. Note that this implementation differs from the work of Ye, *et al.* [56, 77], where they did not need consider this storage issue because they dealt with a two-dimensional problem. The whole optimization procedure is summarized in the pseudo-code of Fig. 4.1.

main {

1. Initialize $\hat{x}$ with a background absorption and diffusion coefficient estimate.
2. Repeat until converged: {
   (a) $\hat{\alpha} \leftarrow \dfrac{1}{P} \parallel y - f(\hat{x}, \hat{s}, \hat{d}) \parallel^2_{\Lambda}$ Eq.(4.17)

   (b) $\hat{s}_k \leftarrow \dfrac{[\ \mathrm{diag}(\hat{d})\ \Phi_k^{(s)}(\hat{x})\ ]^H \Lambda_k^{(s)}\ y}{\parallel \mathrm{diag}(\hat{d})\ \Phi_k^{(s)}(\hat{x})\ \parallel^2_{\Lambda_k^{(s)}}} \qquad k = 1, 2, \ldots, K$ Eq.(4.18)

   (c) $\hat{d}_m \leftarrow \dfrac{[\ \mathrm{diag}(\hat{s})\ \Phi_m^{(d)}(\hat{x})\ ]^H \Lambda_m^{(d)}\ y}{\parallel \mathrm{diag}(\hat{s})\ \Phi_m^{(d)}(\hat{x})\ \parallel^2_{\Lambda_m^{(d)}}} \qquad m = 1, 2, \ldots, M$ Eq.(4.19)

   (d) $\hat{x} \leftarrow ICD_update_x \left\{ c(x, \hat{s}, \hat{d}, \hat{\alpha}), \hat{x} \right\}$ Eq.(4.16)

} }

(a)

$ICD_update_x \left\{ c(x, \hat{s}, \hat{d}, \hat{\alpha}), \hat{x} \right\}$ {

1. Compute $\phi_k(\,\cdot\,; \hat{x})$, $k = 1, 2, \cdots, K$ and $g(b_m,\, \cdot\,; \hat{x})$, $m = 1, 2, \cdots, M$.

2. For $u = 0, 1$,

   For $i = 1, \ldots, N$ (in random order), {

   (a) Compute $[f'(\hat{x}, \hat{s}, \hat{d})]_{*(uN+i)}$ with (4.24)-(4.26).

   (b) Update x_{uN+i}, as described by Ye, *et al.* [77]

$$\begin{aligned} \hat{x}_{uN+i} \quad \leftarrow \quad & \arg \min_{x_{uN+i} \geq 0} \\ & \left\{ \frac{1}{\hat{\alpha}} \left|\left| y - f(\hat{x}, \hat{s}, \hat{d}) - [f'(\hat{x}, \hat{s}, \hat{d})]_{*(uN+i)} \left(x_{uN+i} - \hat{x}_{uN+i} \right) \right|\right|^2_{\Lambda} \right. \\ & \left. + \frac{1}{p_u \sigma^{p_u}} \sum_{j \in \mathcal{N}_i} b_{u,i-j} |x_{uN+i} - \hat{x}_{uN+j}|^{p_u} \right\} \end{aligned}$$ Eq.(4.23)

   }

3. Return $\hat{x}$.

}

(b)

Fig. 4.1. Pseudo-code specification for (a) the overall optimization procedure and (b) the image update by one ICD scan.

4.4 Results

4.4.1 Simulation

The performance of the algorithm described above was investigated by simulation using cubic tissue phantoms of dimension $8 \times 8 \times 8$ cm on an edge and with background $D = 0.03$ cm and $\mu_a = 0.02$ cm^{-1}. Two phantoms were used. Phantom A has two spherical μ_a inhomogeneities with diameters of 2.25 cm and 2.75 cm and central values of 0.070 cm^{-1} that decay smoothly as a fourth order polynomial to the background value, and two spherical D inhomogeneities with diameters of 2.25 cm and a central value of 0.01 cm that increase smoothly to the background value as a fourth order polynomial. Phantom A is shown as an isosurface plot in Fig. 4.2(a,b), and as gray scale plots of cross-sections in Fig. 4.3(a,b). Phantom B has a high absorption inhomogeneity with a peak value of $\mu_a = 0.07$ cm^{-1} near one face of the cube and a low diffusion inhomogeneity near the center with a diameter of 2.75 cm and a central value of 0.01 cm that increases smoothly as a fourth order polynomial to the background value, as shown in Fig. 4.4(a,b) and Fig. 4.5(a,b). Phantom B was used to assess whether an absorber close to a set of sources and detectors is difficult to reconstruct, since its effect might be compensated for by reduced source and detector coupling coefficients. Five sources, with a modulation frequency of 100 MHz, and eight detectors are located on each face (Fig. 4.6a), yielding $K = 30$ and $M = 48$. Shot noise was added to the data, and the average signal-to-noise ratio for sources and detectors on opposite faces was 33 dB. The complex source/detector coupling coefficients (a total of 78 parameters) were generated with a Gaussian distribution centered at $1 + 0i$ and having a standard deviation of $\frac{\sigma_{\text{coeff}}}{\sqrt{2}}(1 + i)$, with $\sigma_{\text{coeff}} = 0.5$ (Fig. 4.7a). The domain was discretized onto $33 \times 33 \times 33$ grid points, and the forward model (4.1) solved using finite differences. Referring to Fig. 4.6(b), a zero-flux ($\phi = 0$) boundary condition on the outer boundary provides the approximate boundary condition on the physical boundary [77, 78]. The sources and detectors were placed 0.6 grid points in from the zero-flux boundary, achieved through appro-

priate weighting of the nearest grid points. Only nodes within the imaging boundary were updated, which excludes the three outermost layers of grid points, to avoid singularities near the sources and detectors. The optimization was initialized using the homogeneous values $D = 0.03$ cm and $\mu_a = 0.02$ cm^{-1}. The image prior model used $p_0 = 2.0$, $\sigma_0 = 0.01$ cm^{-1}, $p_1 = 2.0$, and $\sigma_1 = 0.004$ cm.

Reconstructions of μ_a and D after 30 iterations are shown in Fig. 4.2(c,d) and Fig. 4.3(c,d), for Phantom A, and in Fig. 4.4(c,d) and Fig. 4.5(c,d) for Phantom B. The corresponding images reconstructed with the correct values of coupling coefficients are shown for comparison in Fig. 4.2(e,f), Fig. 4.3(e,f), Fig. 4.4(e,f), and Fig. 4.5(e,f). Our algorithm reconstructs images quite similar to those reconstructed when the true values of the coupling coefficients are used. The corresponding images reconstructed with all coupling coefficients set to $1 + 0i$ are shown in Fig. 4.2(g,h), Fig. 4.4(g,h), Fig. 4.3(g,h) and Fig. 4.5(g,h). These show that poor reconstructions are obtained if the source and detector coupling is not accounted for in the reconstruction process. This is due to the effectively incorrect forward model and hence incorrect Fréchet derivatives. In fact, for the large range of source and detector coupling coefficients used in these examples, the images reconstructed without calibration differ little from the initial starting point of the optimization, when the coupling coefficients are fixed at $1 + 0i$. The convergence of the normalized root mean square error (NRMSE) between the phantoms and the reconstructed images is shown in Fig. 4.8. The NRMSE is defined by

$$\text{NRMSE} \;=\; \left[\frac{1}{2} \sum_{u=0}^{1} \frac{\sum_{r_i \in R} |\hat{x}_{uN+i} - x_{uN+i}|^2}{\sum_{r_i \in R} |x_{uN+i}|^2} \right]^{1/2}, \tag{4.27}$$

where R is the set of the updated grid points within the imaging boundary (shown in Fig. 4.6(b)), $\hat{x}_{uN+i}$ is the reconstructed value of $(uN + i)$-th image element, and x_{uN+i} is the correct value. The NRMSE obtained with the reconstruction incorporating calibration is similar to that obtained when the correct coupling coefficients are used. However, if calibration is not used, there is little decrease in the NRMSE from the starting value.

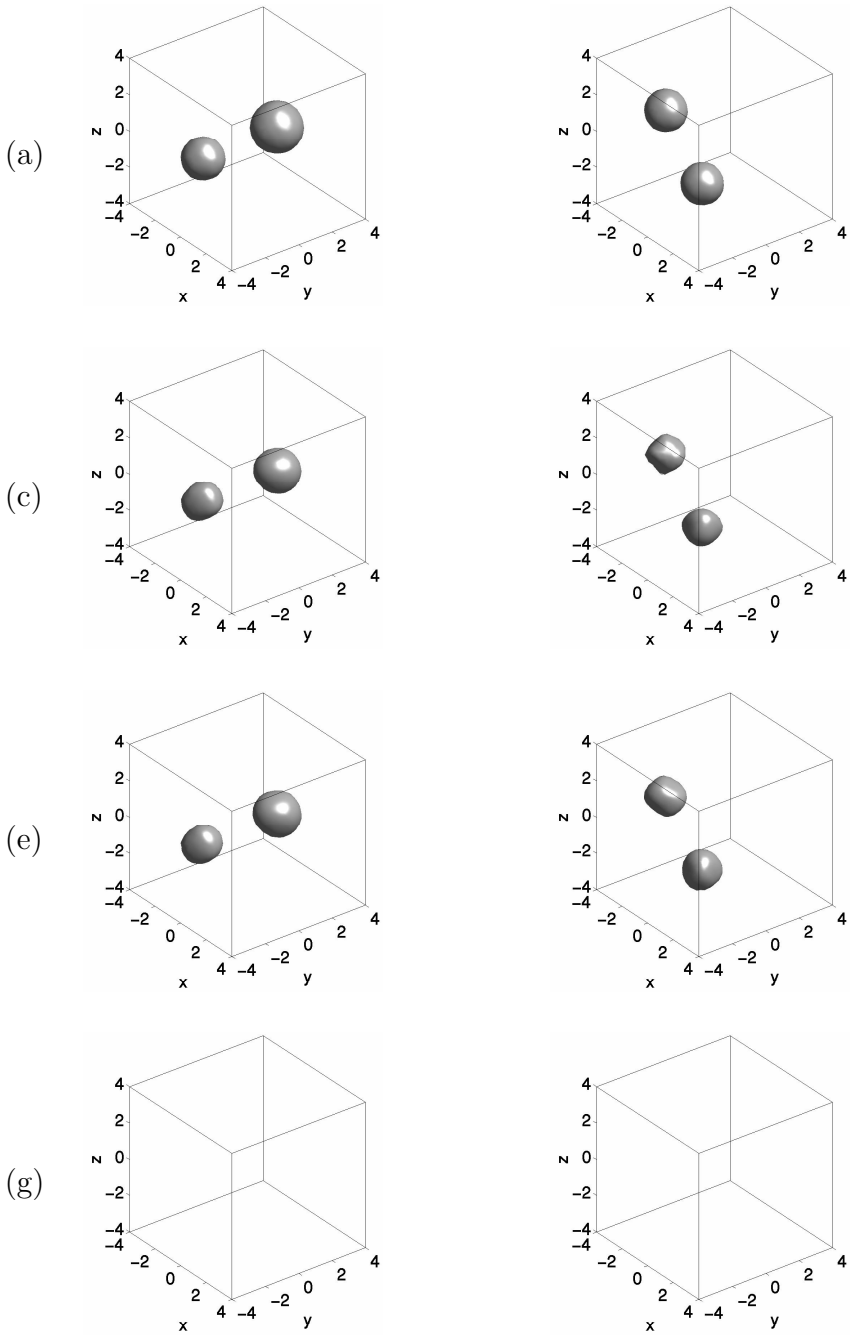


Fig. 4.2. Isosurface plots (at 0.04 cm^{-1} for μ_a, and 0.02 cm for D) for μ_a (left column) and D (right column) for Phantom A: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration.

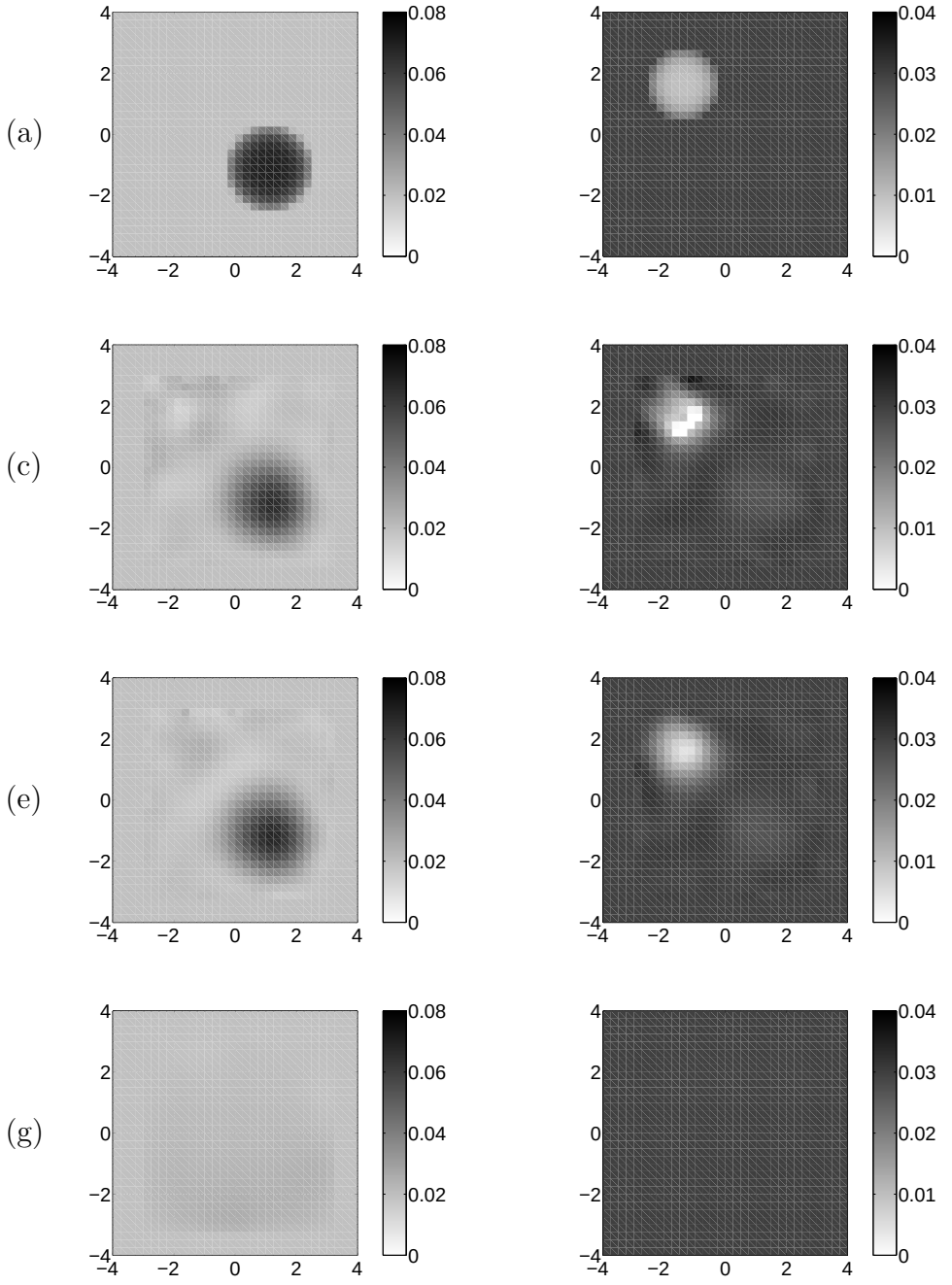


Fig. 4.3. Cross-sections through the centers of the inhomogeneities (z=0.5 cm for μ_a, z=1.5 cm for D) for μ_a (left column) and D (right column) of Phantom A: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration.

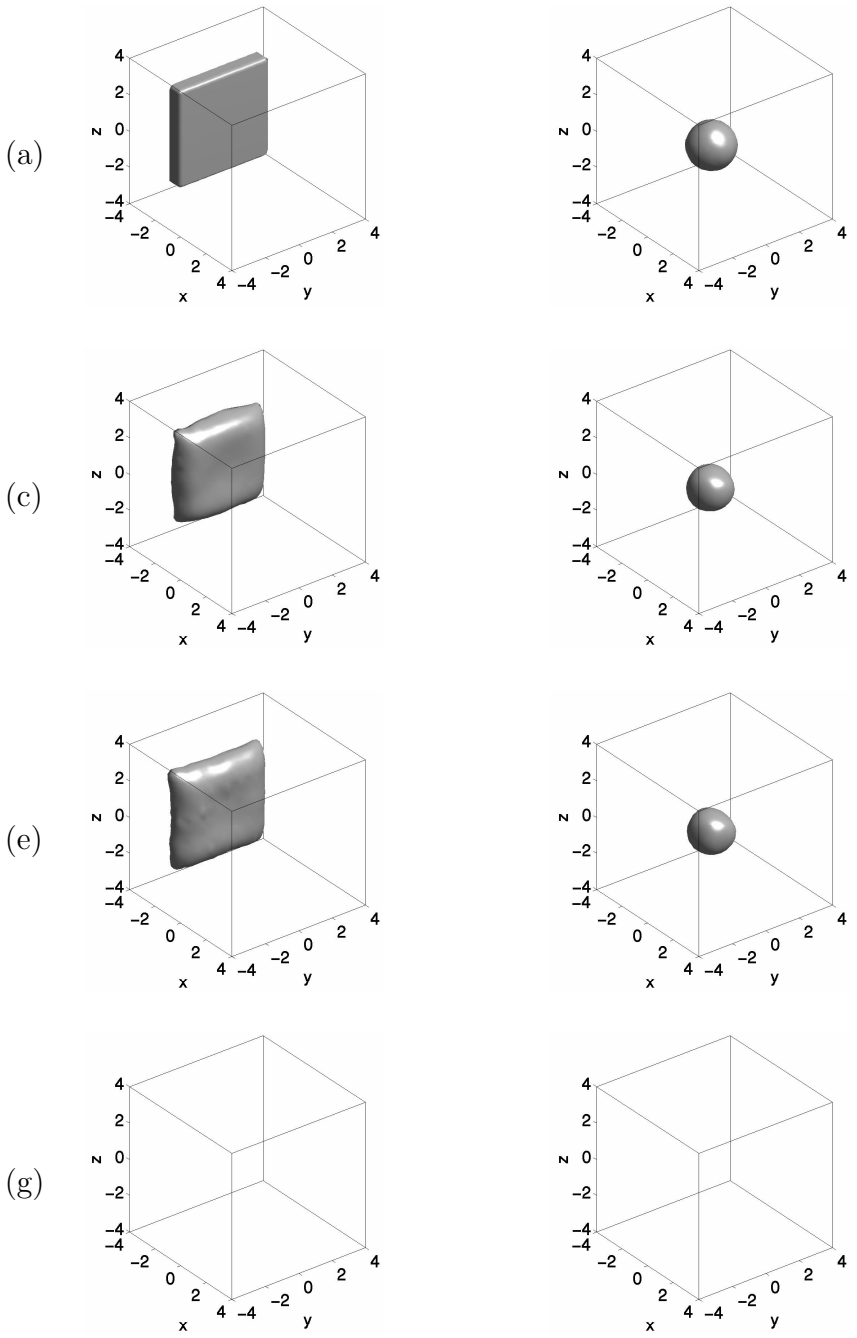


Fig. 4.4. Isosurface plots (at 0.04 cm^{-1} for μ_a, and 0.02 cm for D) for μ_a (left column) and D (right column) for Phantom B: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration.

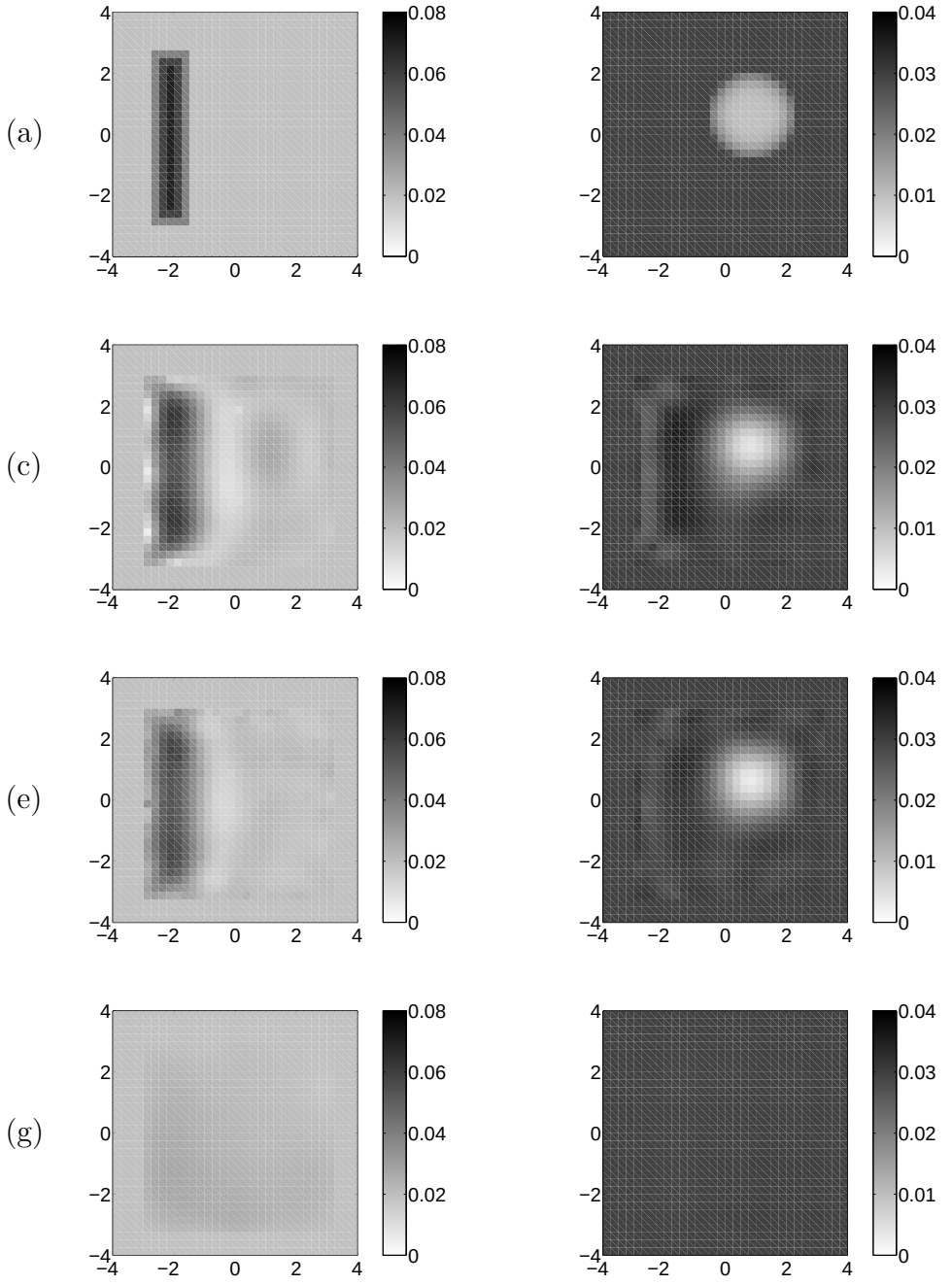
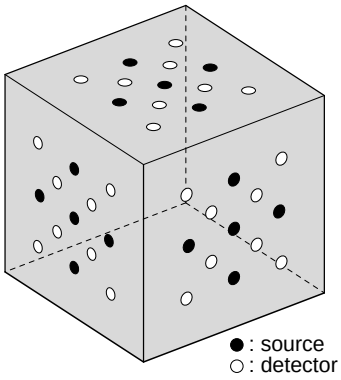
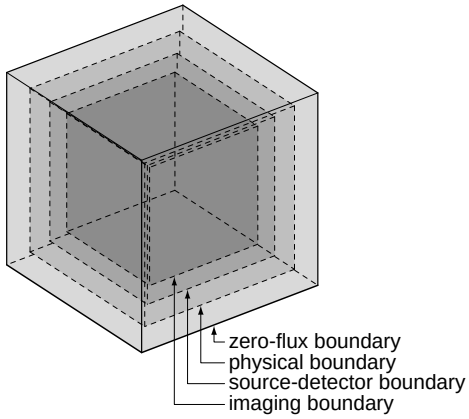


Fig. 4.5. Cross-sections through the centers of the inhomogeneities (z=0.0 cm for μ_a, z=0.25 cm for D) for μ_a (left column) and D (right column) of Phantom B: (a,b) original tissue phantom, (c,d) reconstructions with source-detector calibration, (e,f) reconstructions using the correct weights, (g,h) reconstructions without calibration.



(a)



(b)

Fig. 4.6. (a) Locations of sources and detectors, (b) Several levels of boundaries: zero-flux boundary, physical boundary, source-detector boundary, and imaging boundary, from the outer boundary.

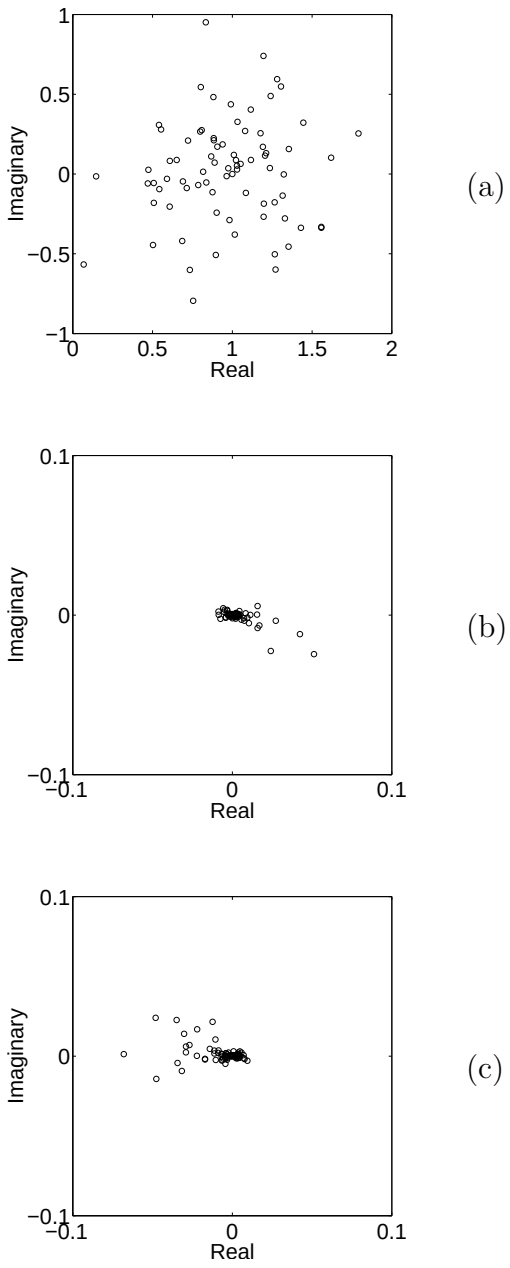
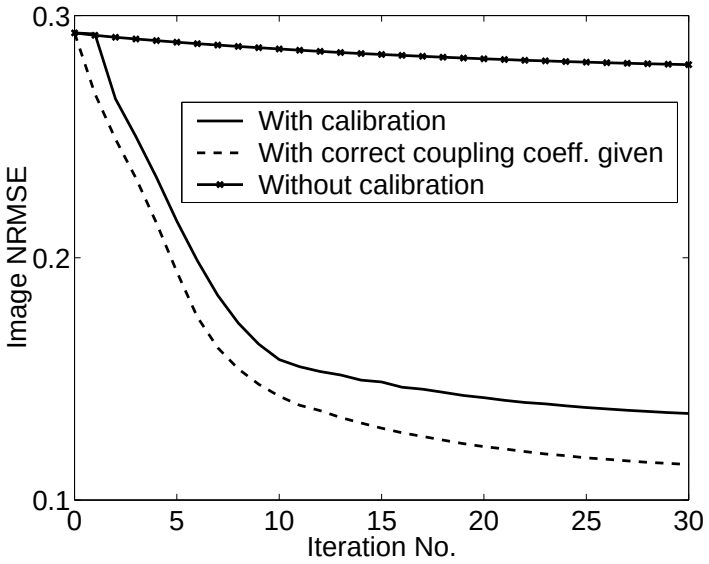
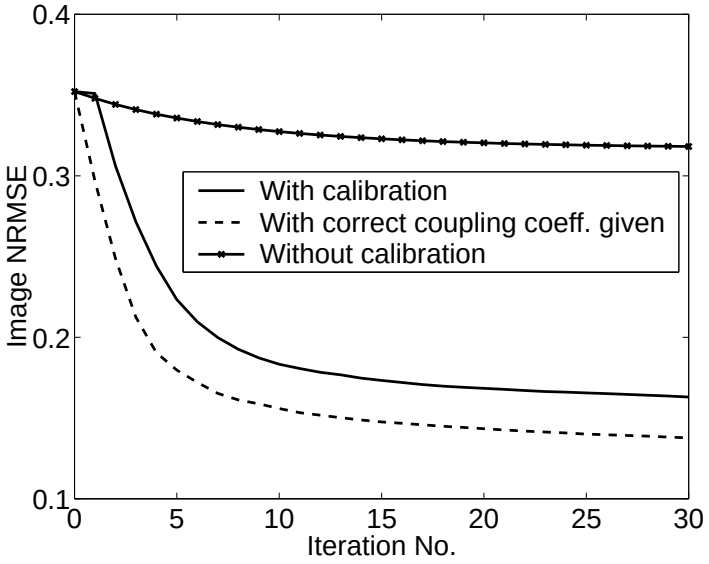


Fig. 4.7. (a) Source/detector coupling coefficients used in the simulations. The estimation error of coupling coefficients for (b) Phantom A and (c) Phantom B after 30 iterations. Note that the scale of (b) and (c) is 10 times of that of (a).

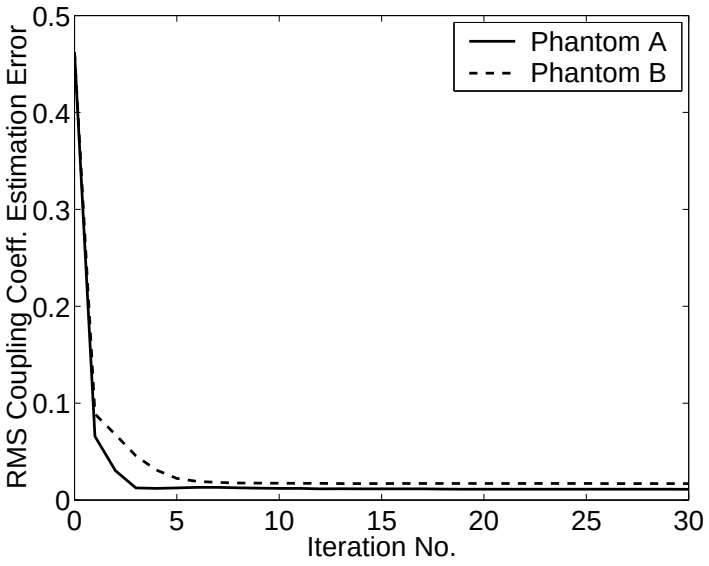


(a)

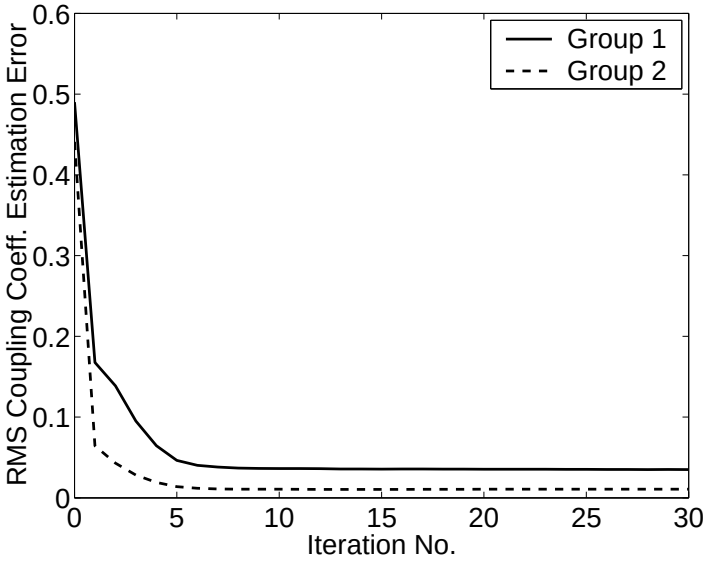


(b)

Fig. 4.8. The normalized root mean square error between the phantom and the reconstructed images for (a) Phantom A and (b) Phantom B.



(a)



(b)

Fig. 4.9. (a) RMS error in the estimated coupling coefficients versus iteration. (b) Convergence of coupling coefficients for Group 1 (—) and Group 2 (- - -) for Phantom B.

The accuracy of the estimated coupling coefficients is shown in Fig. 4.7(b,c), where the differences between the true coupling coefficients and those estimated after 30 iterations is given. The NRMSE error after 30 iterations is 0.011 for Phantom A and 0.017 for Phantom B, which are only 2% and 3% of the standard deviation of the coupling coefficients, respectively, indicating accurate recovery. Figure 4.9 shows the variation of the NRMSE error between the estimated and true coupling coefficients versus iteration, showing good convergence in only a few iterations. The results therefore indicate that our algorithm reconstructs accurate images without prior calibration by the estimation of the coupling coefficients in an efficient optimization scheme.

For Phantom B, the absorber close to one source-detector plane is reconstructed quite accurately and is not distorted by the variable coupling coefficients of the sources and detectors. Some small spikes of low μ_a appear in the neighborhood of some of the sources and detectors (Fig. 4.5(b)), as noted previously [109], but the effect is quite small. However, the final NRMSE is somewhat larger for Phantom B than for Phantom A (Fig. 4.8), and the real part of some of the coupling coefficients is underestimated (Fig. 4.7(c)). We categorize the sources and detectors on the side nearest the absorber as Group 1, and the remainder as Group 2. Most of the underestimated coefficients are those for sources and detectors on the face close to the absorber. The estimation error for these coupling coefficients (Group 1) is larger than the remaining sources and detectors (Fig. 4.9(b)). Therefore, because the light transmitted through the absorber is highly attenuated, it is partially compensated for by reduced estimated coupling coefficients. As noted above, however, the effect is quite small.

In order to study the effect of the variability of the coupling coefficients, reconstructions were performed for Phantom A for different standard deviations of the (real and imaginary parts of the) coupling coefficients, σ_{coeff}. The coupling coefficients were generated with a Gaussian distribution centered at $1 + 0i$ and having $\frac{\sigma_{coeff}}{\sqrt{2}}(1+i)$, and images are the reconstructed results after 30 iterations of our algo-

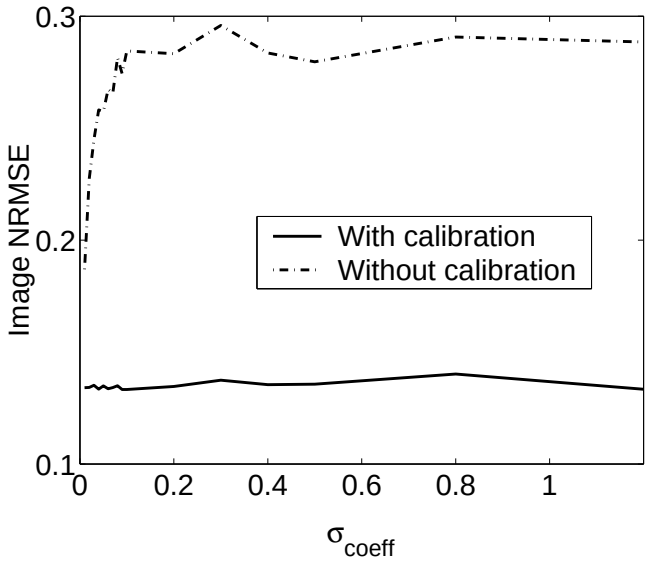


Fig. 4.10. Image NRMSE comparison between the reconstruction with coupling coefficient calibration and the reconstruction with coupling coefficients fixed to $1 + 0i$, for various standard deviations of coupling coefficients. Images were obtained after 30 iterations.

rithm. The image NRMSE is compared for various standard derivations in Fig. 4.10. Estimating the calibration coefficients reduces the NRMSE, as expected. The error without calibration did not increase beyond about 0.28 with increasing σ_{coeff}, as this value for the image NRMSE corresponds to the initial value with the correct background parameters and indicates that an image is not recovered. To establish the gradual deterioration of the image with source-detector coupling coefficients that are not accounted for in the reconstruction, Fig. 4.11(a,b) shows the image obtained with for $\sigma_{\text{coeff}} = 0.02$ and Fig. 4.11(c,d) that for $\sigma_{\text{coeff}} = 0.04$, as compared with the true images in Fig. 4.3(a,b). This result indicates that accurate estimation of the coupling coefficients is crucial for determining accurate images. The σ_{coeff} will obviously be a function of the specific experimental arrangement. Figure 4.10 serves as an illustration of the impact of variations in the source-detector coupling. While some experimental arrangements may have (approximately) a single, scalar source-detector weight [13], it is still important to determine this value.

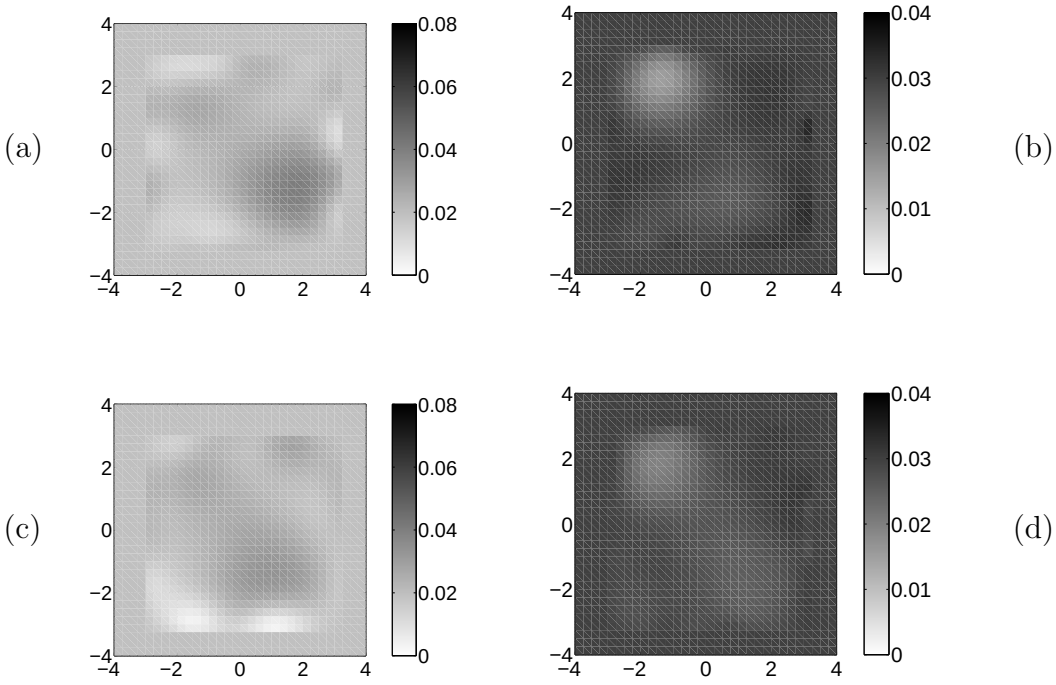


Fig. 4.11. Cross-sections of the reconstructed images through the centers of the inhomogeneities (z=0.5 cm for μ_a, z=1.5 cm for D) : for $\sigma_{\text{coeff}} = 0.02$ for (a) μ_a and (b) D, and for $\sigma_{\text{coeff}} = 0.04$ for (c) μ_a and (d) D.

We have established that multi-resolution techniques such as multigrid achieve more reliable convergence of the cost function while dramatically reducing the computation time in two-dimensional optical diffusion tomography. [56] The approach presented for extracting the source-detector weights as part of the image reconstruction in a Bayesian framework could be extended to multi-resolution approaches. We investigated a simple multi-resolution approach by using a coarse grid solution ($17 \times 17 \times 17$) to initialize a fine grid solution ($33 \times 33 \times 33$). Better convergence was achieved using this simple two-grid approach with various initial conditions consisting of uniform D and μ_a differing from the true background by as much as a factor of three. This performance improvement occurs both with known and estimated source-detector weights. Also, we noticed that in some cases with a fixed, fine grid, the cost function with variable source-detector weights was slightly larger than that with the true weights set. While the images in these cases were still excellent, the additional degrees of freedom should have resulted in a smaller value of the cost function. Using the multi-resolution approach, this was indeed the case, providing further evidence of the robustness of our approach. We emphasize that the algorithm we present for extraction of the source-detector weights in a Bayesian framework was consistently effective, regardless of the particular iterative reconstruction approach.

4.4.2 Experiment

The effectiveness of our source-detector calibration approach was evaluated for measurements made on an optically clear culture flask containing a black plastic cylinder embedded in a turbid suspension (Fig. 4.12(a)). The plastic cylinder was embedded in a 0.5% concentration Intralipid solution. The data was collected with an inexpensive apparatus comprised of an infrared LED operating at 890 nm and a silicon p-i-n photodiode, as schematically depicted in Fig. 4.12(b). With the source centrally located, as shown in Fig. 4.12(b), the detector located on the other side of the flask was mechanically scanned in the same plane as the source, and data

were taken at 25 symmetrical locations. The flask was rotated, so that the relative positions of source and detector were reversed, and another set of data taken. This resulted in a total of two source positions with 25 detector measurements each. The sources were modulated at 50 MHz. This experimental arrangement is similar to one we used previously [6, 13], but with two sources instead of one.

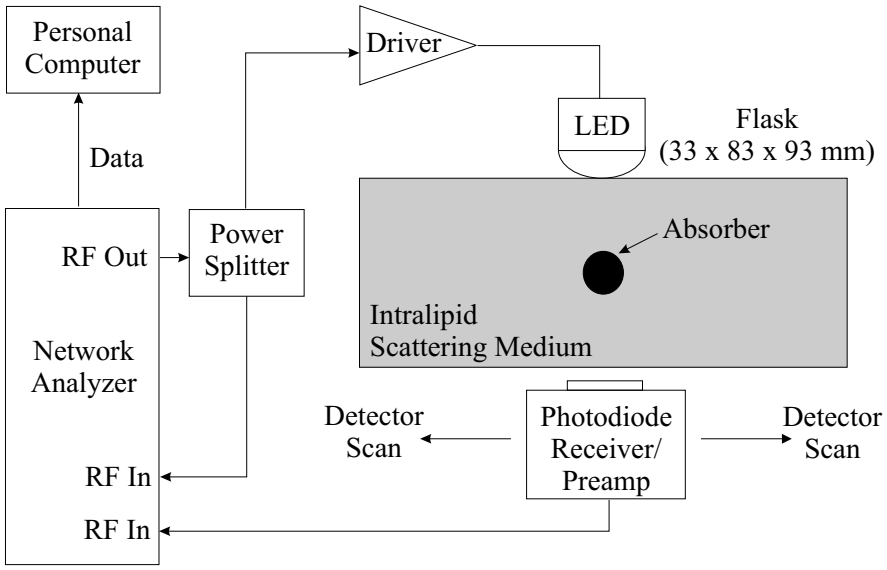
For this experiment, each set of 25 measurements used a single detector that was translated, so we modeled all 25 measurements with a single detector calibration parameter. In addition, there are two source calibration parameters. Without loss of generality, however, the two source calibration parameters were assumed to be 1 since, for this experiment, any change in source phase and amplitude can be equivalently accounted for by the detector calibration parameters. Therefore, a total of two unknown calibration parameters, i.e., two detector calibration parameters, were estimated.

Inversions were performed for the absorption coefficients and coupling coefficients, assuming D known. The domain was discretized into $65 \times 33 \times 65$ grid points. For computational efficiency, we used a simple multiresolution technique in which 200 coarse grid ($33 \times 17 \times 33$) iterations are followed by 30 fine grid iterations. We used $\sigma_0 = 1.0$ cm^{-1} and $p_0 = 2.0$ for the image prior model.

Figure 4.13 contains reconstructed images of the absorption coefficient in the measurement plane. Figure 4.13(a) shows the reconstruction obtained using two complex valued calibration coefficients; Figure 4.13(b) shows the reconstruction obtained when only a single complex calibration coefficient was used (i.e., the two coefficients were assumed equal); Figure 4.13(c) shows the reconstruction obtained with a single real valued calibration coefficient; and finally, Figure 4.13(d) assumed all calibration coefficients to be 1. The reconstruction of Fig. 4.13(a) used the most accurate model and also produced a reconstruction that appears to be most accurate in shape. Because we used the same type of sources, the difference between two source calibration coefficients was not significant. Therefore, Fig. 4.13(b) shows almost the same reconstruction quality as Fig. 4.13(a), but with slightly more arti-



(a)



(b)

Fig. 4.12. (a) Culture flask with the absorbing cylinder embedded in a scattering Intralipid solution. (b) Schematic diagram of the apparatus used to collect data.

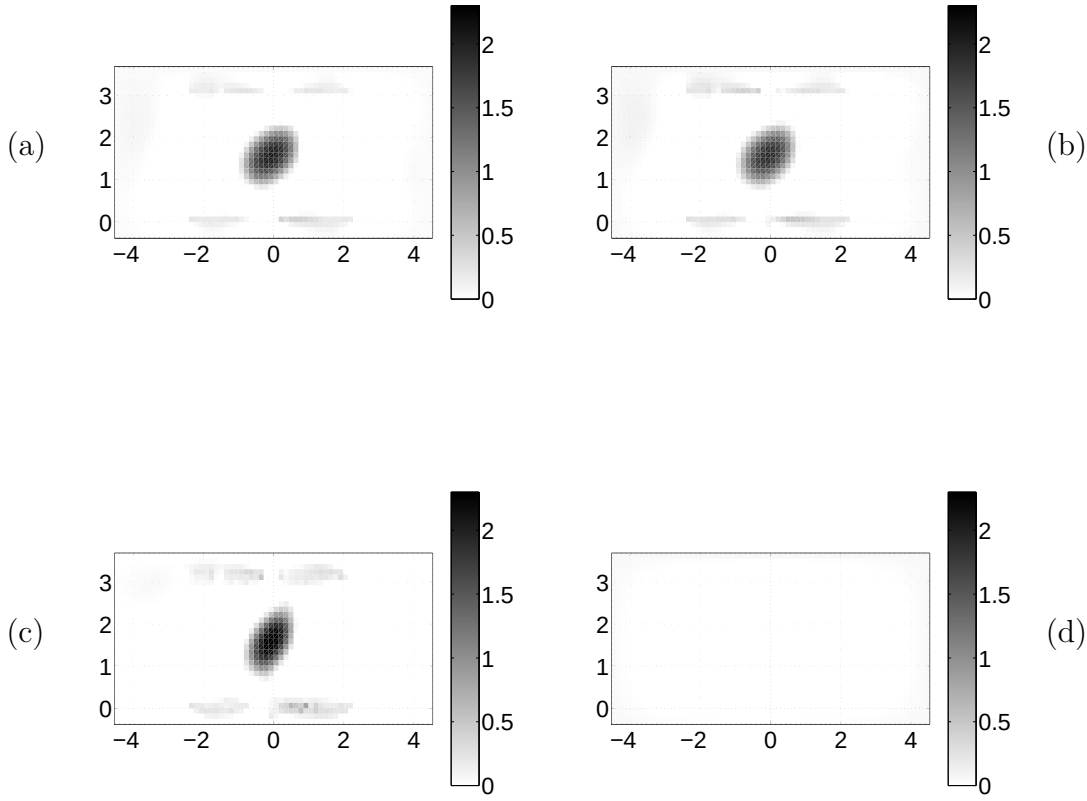


Fig. 4.13. Cross-sections for reconstructed images of an absorbing cylinder with (a) two complex valued calibration coefficients, (b) a single complex calibration coefficient, (c) a single real calibration coefficient, and (d) all calibration coefficients assumed to be 1.

facts in the neighborhood of the detector locations. Generally, the elliptical shape of the reconstruction in Fig. 4.13(c) appears to be the least accurate. Importantly, Fig. 4.13(d) shows that reconstruction without accurate estimation of the calibration coefficients was not possible.

4.5 Conclusions

We have formulated the Bayesian optical diffusion tomography with the source-detector parameter estimation problem and proposed an efficient optimization scheme. Our algorithm does not require any prior calibration, and it estimates coupling coefficients successfully with only a small amount of additional computation. Simulation and experimental results show that images can be reconstructed along with the accurate estimation of the coupling coefficients.

LIST OF REFERENCES

LIST OF REFERENCES

[1] D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang. Imaging the body with diffuse optical tomography. *IEEE Signal Proc. Magazine*, 18(6):57–75, Nov. 2001.

[2] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15:R41–R93, 1999.

[3] J. C. Hebden, S. R. Arridge, and D. T. Delpy. Optical imaging in medicine: I. experimental techniques. *Phys. Med. Biol.*, 42:825–840, 1997.

[4] S. R. Arridge and J. C. Hebden. Optical imaging in medicine: II. Modelling and reconstruction. *Phys. Med. Biol.*, 42:825–840, 1997.

[5] G. J. Saulnier, R. S. Blue, J. C. Newell, D. Isaacson, and P. M. Edic. Electrical impedance tomography. *IEEE Signal Proc. Magazine*, 18(6):31–43, Nov. 2001.

[6] J. S. Reynolds, A. Przadka, S. Yeung, and K. J. Webb. Optical diffusion imaging: a comparative numerical and experimental study. *Applied Optics*, 35(19):3671–3679, July 1996.

[7] E. C. Fear, S. C. Hagness, P. M. Meaney, M. Okoniewski, and M. A. Stuchly. Enhancing breast tumor detection with near field imaging. *IEEE Microwave Magazine*, pages 48–56, March 2002.

[8] R. L. Thomas. Reflections of a thermal wave imager: tow decades of research in photoacoustics and photothermal phenomena. *Analytical Sciences*, 17:s1–s4, April 2001.

[9] R. Pike and P. Sabatier. *Scattering: Scattering and Inverse Scattering in Pure and Applied Science*. Academic Press, San Diego, 2002.

[10] K. Sauer and C. A. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. on Signal Processing*, 41(2):534–548, February 1993.

[11] M. V. Ranganath, A. P. Dhawan, and N. Mullani. A multigrid expectation maximization reconstruction algorithm for positron emission tomography. *IEEE Trans. on Medical Imaging*, 7(4):273–278, Dec. 1988.

[12] T. Pan and A. E. Yagle. Numerical study of multigrid implementations of some iterative image reconstruction algorithms. *IEEE Trans. on Medical Imaging*, 10(4):572–588, Dec. 1991.

[13] A. B. Milstein, S. Oh, J. S. Reynolds, K. J. Webb, C. A. Bouman, and R. P. Millane. Three-dimensional Bayesian optical diffusion tomography using experimental data. *Optics Letters*, 27:95–97, January 2002.

[14] S. Oh, A. B. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb. Source-detector calibration in three-dimensional Bayesian optical diffusion tomography. *J. Optical Society America A*, 19(10):1983–1993, Oct. 2002.

[15] A. B. Milstein, S. Oh, K. J. Webb, C. A. Bouman, Q. Zhang, D. A. Boas, and R. P. Millane. Fluorescence optical diffusion tomography. *Applied Optics*, *to appear*.

[16] B. Sahiner and A. Yagle. Image reconstruction from projections under wavelet constraints. *IEEE Trans. on Signal Processing*, 41(12):3579–3584, 1993.

[17] M. Bhatia, W. C. Karl, and A. S. Willsky. Wavelet-based method for multiscale tomographic reconstruction. *IEEE Trans. on Medical Imaging*, 15(1):92–101, 1996.

[18] M. Bhatia, W. C. Karl, and A. S. Willsky. Tomographic reconstruction and estimation based on multiscale natural -pixel bases. *IEEE Trans. on Image Processing*, 6(3):463–478, March 1997.

[19] N. Lee. *Wavelet-vaguelette decompositions and homogenous equations*. Ph.D. dissertation, Purdue University, West Lafayette, IN, 1998.

[20] A. Delaney and Y. Bresler. Multiresolution tomographic reconstruction using wavelets. *IEEE Trans. on Image Processing*, 4(6):799–813, June 1995.

[21] Z. Wu, G. T. Herman, and J. A. Browne. Edge preserving reconstruction using adaptive smoothing in wavelet domain. In *Proc. of IEEE Nucl. Sci. Symp. and Med. Imaging Conf.*, volume 3, pages 1917–1921, San Francisco, CA, October 31 - November 6 1993.

[22] S. S. Saquib, C. A. Bouman, and K. Sauer. A non-homogeneous MRF model for multiresolution Bayesian estimation. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume 2, pages 445–448, Lausanne Switzerland, September 16-19 1996.

[23] R. Nowak and E. D. Kolaczyk. A multiscale MAP estimation method for Poisson inverse problems. In *Proceedings of the 32nd Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1682–1686, Pacific Grove, CA, November 1-4 1998.

[24] T. Frese, C. A. Bouman, and K. Sauer. Adaptive wavelet graph model for Bayesian tomographic reconstruction. *IEEE Trans. on Image Processing*, 11(7):756–770, July 2002.

[25] E. L. Miller, L. Nicolaides, and A. Mandelis. Nonlinear inverse scattering methods for thermal wave slice tomography: A wavelet domain approach. *J. Optical Society America A*, 15(6):1545–1556, June 1998.

[26] W. Zhu, Y. Wang, Y. Deng, Y. Yao, and R. Barbour. A wavelet-based multiresolution regularization least squares reconstruction approach for optical tomography. *IEEE Trans. on Medical Imaging*, 16(2):210–217, April 1997.

[27] A. Brandt. Multi-level adaptive solutions to boundary value problems. *Mathematics of Computation*, 31(138):333–390, April 1977.

[28] U. Trottenberg, C.W. Oosterlee, and A. Schueller. *Multigrid*. Academic Press, London, 2000.

[29] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial, 2nd Ed.* Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[30] S. McCormick, editor. *Multigrid Methods*. Society for Industrial and Applied Mathematics, Philadelphia, 1987.

[31] W. Hackbusch. *Multigrid Methods and Applications*. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1985.

[32] P. Wesseling. *An Introduction to Multigrid Methods*. John Wiley & Sons, Chichester, 1992.

[33] D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(2):129–139, March 1986.

[34] R. Kimmel and I. Yavneh. An algebraic multigrid approach for image analysis. *SIAM Journal of Scientific Computing*, 24(4):1218–1231, 2003.

[35] E. Enkelmann. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Comput. Vision Graphics and Image Process.*, 43:150–177, 1988.

[36] E. Memin and P. Perez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. on Image Processing*, 7(5):703–719, May 1998.

[37] P. Hellier, C. Barillot, E. Mémin, and P. Pérez. Hierarchical estimation of a dense deformation field for 3-d robust registration. *IEEE Trans. on Medical Imaging*, 20(5):388–402, May 2001.

[38] S. Ghosal and P. Vanek. Fast algebraic multigrid for discontinuous optical flow estimation. Technical Report UCD-CCM-025, Center for Computational Mathematics, University of Colorado at Denver, 1994.

[39] P. Saint-Marc, J. Chen, and G. Medioni. Adaptive smoothing: a general tool for early vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(6):514–529, June 1991.

[40] M. Unser. Multigrid adaptive image processing. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume I, pages 49–52, Washington DC, USA, Oct. 1995.

[41] D. L. Pham and J. L. Prince. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans. on Medical Imaging*, 18(9):737–752, Sep. 1999.

[42] S. Henn and K. Witsch. A multigrid approach for minimizing a nonlinear functional for digitial image matching. *Computing*, 64:339–348, 2000.

[43] K. Zhou and C. K. Rushforth. Image restoration using multigrid methods. *Applied Optics*, 30(20):2906–2912, July 1991.

[44] S. T. Acton. Multigrid anisotropic diffusion. *IEEE Trans. on Image Processing*, 7(3):280–291, March 1998.

[45] D. Terzopoulos. The computation of visible-surface representations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(4):417–438, July 1988.

[46] M. Arigovindan, M. Sühling, P. Hunziker, and M. Unser. Multigrid image reconstruction from arbitrarily spaced samples. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume III, pages 381–384, Rochester NY, USA, Sep. 22-25, 2002.

[47] C. A. Bouman and K. Sauer. Nonlinear multigrid methods of optimization in Bayesian tomographic image reconstruction. In *Proc. of SPIE Conf. on Neural and Stochastic Methods in Image and Signal Processing*, volume 1766, pages 296–306, San Diego, CA, July 19-24 1992.

[48] S. F. McCormick and J. G. Wade. Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography. *Inverse Problems*, 9:697–713, 1993.

[49] L. Borcea. Nonlinear multigrid for imaging electrical conductivity and permittivity at low frequency. *Inverse Problems*, 17:329–359, April 2001.

[50] R. Gandlin and A. Brandt. Two multigrid algorithms for inverse problem in electrical impedance tomography. In *Proc. 2003 Copper Mountain Conf. Multigrid Methods*, Copper Mountain, CO, USA, March 30-April 4 2003.

[51] A. Brandt and R. Gandlin. Multigrid for atmospheric data assimilation: analysis. In *Proc. 2002 Hyperbolic Problems: Theory, Numerics and Applications*, pages 369–376, Pasadena, CA, USA, March 2002.

[52] A. Brandt. Multiscale and multiresolution methods: Theory and applications. In T. J. Barth, T. F. Chan, , and R. Haimes, editors, *Multiscale Scientific Computation: Review 2001*, pages 3–96. Springer Verlag, Heidelberg, 2001.

[53] A. Brandt and D. Ron. Multigrid solvers and multilevel optimization strategies. In J. Cong and J.R. Shinnerl, editors, *Multilevel Optimization and VLSICAD*, pages 1–69. Kluwer Academic Publishers, Boston, 2002.

[54] C. R. Johnson, M. Mohr, U. Ruede, A. Samsonov, and K. Zyp. Multilevel methods for inverse bioelelectric field problems. In *Lecture Notes in Computational Science and Engineering - Multiscale and Multiresolution Methods: Theory and Applications, Eds. T.J. Barth, T.F. Chan, R. Haimes.*, volume 20, Springer-Verlag Publishing, Heidelberg, Oct 2001.

[55] J. C. Ye, C. A. Bouman, R. P. Millane, and K. J. Webb. Nonlinear multigrid optimization for Bayesian diffusion tomography. In *Proc. of IEEE Int'l Conf. on Image Proc.*, Kobe, Japan, October 25-28 1999.

[56] J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane. Nonlinear multigrid algorithms for Bayesian optical diffusion tomography. *IEEE Trans. on Image Processing*, 10(6):909–922, June 2001.

[57] S. G. Nash. A multigrid approach to discretized optimization problems. *J. of Optimization methods and software*, 14:99–116, 2000.

[58] R. M. Lewis and S. G. Nash. A multigrid approach to the optimization of systems governed by differential equations. In *8-th AIAA/USAF/ISSMO Symp. Multidisciplinary Analysis and Optimization*, Long Beach, CA, 2000.

[59] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. Multigrid inversion algorithms with applications to optical diffusion tomography. In *Proc. of 36th Asilomar Conference on Signals, Systems, and Computers*, pages 901–905, Monterey, CA, Nov. 2002.

[60] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. Multigrid algorithms for optimizations and inverse problems. In *2003 Electronic Imaging*, Santa Clara, CA, USA, Jan. 20-25 2003.

[61] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. Adaptive nonlinear multigrid inversion with applications to Bayesian optical diffusion tomography. In *Proc. IEEE Workshop on Statistical Signal Processing*, St. Louis, MO, USA, Sep. 2003.

[62] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. Nonlinear multigrid inversion. In *Proc. of IEEE Int'l Conf. on Image Proc.*, Barcelona, Spain, Sep. 2003.

[63] S. S. Saquib, K. M. Hanson, and G. S. Cunningham. Model-based image reconstruction from time-resolved diffusion data. In *Proc. of SPIE Conf. on Medical Imaging: Image Processing*, volume 3034, pages 369–380, Newport Beach, CA, February 25-28 1997.

[64] A. H. Hielscher, A. D. Klose, and K. M. Hanson. Gradient-based iterative image reconstruction scheme for time-resolved optical tomography. *IEEE Trans. on Medical Imaging*, 18(3), March 1999.

[65] A. Mohammad-Djafari. Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems. In *Proc. of IEEE Int'l Conf. on Image Proc.*, volume II, pages 473–476, Lausanne, Switzerland, September 16-19 1996.

[66] A. Mohammad-Djafari. On the estimation of hyperparameters in Bayesian approach of solving inverse problems. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pages 495–498, Minneapolis, Minnesota, April 27-30 1993.

[67] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statistics*, 37:1554–1563, 1966.

[68] L. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statistics*, 41(1):164–171, 1970.

[69] A.R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. on Medical Imaging*, 14(1):132–137, March 1995.

[70] J. Browne and A. R. De Pierro. A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography. *IEEE Trans. on Medical Imaging*, 15(5):687–699, October 1996.

[71] J.A. Fessler, E.P. Ficaro, N.H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmi ssion image reconstruction. *IEEE Trans. on Medical Imaging*, 16(2):166–175, April 1997.

[72] J. Zheng, S. S. Saquib, K. Sauer, and C. A. Bouman. Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence. *IEEE Trans. on Image Processing*, 9(10):1745–1759, Oct. 2000.

[73] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36(2):192–236, 1974.

[74] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, 8(2):194–202, June 1989.

[75] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992.

[76] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. on Image Processing*, 2(3):296–310, July 1993.

[77] J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane. Optical diffusion tomography using iterative coordinate descent optimization in a Bayesian framework. *J. Optical Society America A*, 16(10):2400–2412, October 1999.

[78] J. C. Ye, K. J. Webb, R. P. Millane, and T. J. Downar. Modified distorted Born iterative method with an approximate Fréchet derivative for optical diffusion tomography. *J. Optical Society America A*, 16(7):1814–1826, July 1999.

[79] J. C. Adams. MUDPACK: Multigrid portable FORTRAN software for the efficient solution of linear elliptic partial differential equations. *Appl. Math. Comput.*, 34:113–146, 1989.

[80] Z. Kato, M. Berthod, and J. Zerubia. Parallel image classification using multiscale Markov random fields. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, volume 5, pages 137–140, Minneapolis, MN, April 27-30 1993.

[81] C. A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2):162–177, March 1994.

[82] M. L. Comer and E. J. Delp. Segmentation of textured images using a multiresolution Gaussian autoregressive model. *IEEE Trans. on Image Processing*, 8(3):408–420, March 1999.

[83] J-M. Laferte, P. Perez, and F. Heitz. Discrete Markov image modeling and inference on the quadtree. *IEEE Trans. on Image Processing*, 9(3):390–404, March 2000.

[84] R. D. Nowak. Shift invariant wavelet-based statistical models and 1/f processes. In *IEEE DSP Workshop*, 1998.

[85] K. Chou, A. Willsky, A. Benveniste, and M. Basseville. Recursive and iterative estimation algorithms for multi-resolution stochastic processes. In *Proceedings of the 28^{t}h Conference on Decision and Control*, volume 2, pages 1184–1189, Tampa, Florida, December 13-15 1989.

[86] H-C. Yang and R. Wilson. Adaptive image restoration using a multiresolution Hopfield neural network. In *Fifth International conference on Image Processing and its applications, (IEE Conference Publication No.410)*, pages 198–202, Edinburgh, UK, July 4-6 1995.

[87] R. D. Nowak and E. D. Kolaczyk. A statistical multiscale framework for Poisson inverse problems. *IEEE Trans. on Information Theory*, 46(5):1811–1825, August 2000. Special Issue on Information-Theoretic Imaging.

[88] G. Wang, J. Zhang, and G. Pan. Solution of inverse problems in image processing by wavelet expansion. *IEEE Trans. on Image Processing*, 4(5):579–591, May 1995.

[89] N. Lee and B. J. lucier. Wavelet methods for inverting the radon transform with noisy data. *IEEE Trans. on Image Processing*, 10(1):79–94, Jan. 2001.

[90] V. E. Henson, M. A. Limber, S. F. McCormick, and B. T. Robinson. Multilevel image reconstruction with natural pixels. *SIAM J. Sci. Comp.*, 17:193–216, 1996.

[91] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. A general framework for nonlinear multigrid inversion. *IEEE Trans. on Image Processing*, 14(1):125–140, Jan 2005.

[92] J. A. O'Sullivan and J. Benac. Alternating minimization multigrid algorithms for transmission tomography. In *Proc. of SPIE Conf. on Computational Imaging II*, pages 216–21, San Jose, California, USA, Jan. 2004.

[93] T. Olson and J. DeStefano. Wavelet localization of the Radon transform. *IEEE Trans. on Signal Processing*, 42:2055–2067, 1994.

[94] T. Olson. Optimal time-frequency projections for localized tomography. *Ann. Biomed. Eng.*, 23:622–636, 1995.

[95] B. Sahiner and A. Yagle. Region-of-interest tomography using exponential radial sampling. *IEEE Trans. on Image Processing*, 4(8):1120–1127, August 1995.

[96] R. Rashid-Farrokhi, K. J. R. Liu, C. A. Berenstein, and D. Walnut. Wavelet-based multiresolution local tomography. *IEEE Trans. on Image Processing*, 6:1412–1430, 1997.

[97] S. Y. Zhao, G. Welland, and G. Wang. Wavelet sampling and localization schemes for the Radon transform in two dimensions. *SIAM J. Appl. Math.*, 57:1749–1762, 1997.

[98] S. Y. Zhao. Wavelet filtering for filtered backprojection in computed tomography. *Appl. Comput. Harmon. Anal.*, 6:346–373, 1999.

[99] A. Brandt, J. Mann, M. Brodski, and M. Galun. A fast and accurate multilevel inversion of the radon transform. *SIAM J. Appl. Math.*, 60(2):437–462, 1999.

[100] H.M. Hudson and R.S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. on Medical Imaging*, 13(4):601–609, December 1994.

[101] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. on Image Processing*, 5(3):480–492, March 1996.

[102] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, November 1984.

[103] E. Levitan and G. Herman. A maximum a *posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. on Medical Imaging*, MI-6:185–192, September 1987.

[104] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. on Nuclear Science*, 23:1428–1432, 1976.

[105] A. J. Rockmore and A. Macovski. A maximum likelihood approach to transmission image reconstruction from projections. *IEEE Trans. on Nuclear Science*, 234:1929–1935, 1977.

[106] S. F. McCormick. *Multilevel adaptive methods for partial differential equations.* SIAM, Philadelphia, 1989.

[107] S. A. Shepp and B. F. Logan. The Fourier reconstruction of a head section. *IEEE Trans. on Nuclear Science*, NS-21:21–43, 1974.

[108] S. R. Arridge and M. Schweiger. A gradient-based optimisation scheme for optical tomography. *Optics Express*, 2(6):213–226, March 1998.

[109] D. Boas, T. Gaudette, and S. Arridge. Simultaneous imaging and optode calibration with diffuse optical tomography. *Opt. Express*, 8(5):263–270, February 2001.

[110] H. Jiang, K. Paulsen, and U. Osterberg. Optical image reconstruction using dc data: simulations and experiments. *Phys. Med. Biol.*, 41(8):1483–1498, August 1996.

[111] H. Jiang, K. Paulsen, U. Osterberg, and M. Patterson. Improved continuous light diffusion imaging in single- and multi-target tissue-like phantoms. *Phys. Med. Biol.*, 43(3):675–693, March 1998.

[112] B. W. Pogue, S. P. Poplack, T. O. McBride, W. A. Wells, K. S. Osterman, U. L. Osterberg, and K. D. Paulsen. Quantitavie hemoglobin tomography with diffuse near-infrared spectroscopy: pilot results in the breast. *Radiology*, 218:261–266, Jan. 2001.

[113] B. W. Pogue, C. Willscher, T. O. McBride, U. L. Osterberg, and K. D. Paulsen. Constrast-detail analysis for detection and characterization with near-infrared diffuse tomography. *Med. Phys.*, 27(12):2693–2700, Dec. 2000.

[114] T. O. McBride, B. W. Pogue, S. Poplack, S. Soho, W. A. Wells, S. Jiang, U. Osterberg, and K. D. Paulsen. Multispectral near-infrared tomography: a case study in compensating for water and lipid content in hemoglobin imaging of the breast. *Journal of Biomed. Optics*, 7(1):72–79, Jan. 2002.

[115] N. Iftimia and H. Jiang. Quantitative optical image reconstructions of turbid media by use of direct-current measurements. *Applied Optics*, 39(28):5256–5261, October 2000.

[116] J. J. Duderstadt and L. J. Hamilton. *Nuclear Reactor Analysis*. Wiley, New York, 1976.

[117] A. Ishimaru. *Wave Propagation and Scattering in Random Media*, volume 1. Academic Press, New York, New York, 1978.

[118] S. Geman and D. McClure. Statistical methods for tomographic image reconstruction. *Bull. Int. Stat. Inst.*, LII-4:5–21, 1987.

[119] S. S. Saquib, C. A. Bouman, and K. Sauer. ML parameter estimation for Markov random fields with applications to Bayesian tomography. *IEEE Trans. on Image Processing*, 7(7):1029–1044, July 1998.

[120] K. Lange. An overview of Bayesian methods in image reconstruction. In *Proc. of the SPIE Conference on Digital Image Synthesis and Inverse Optics*, volume SPIE-1351, pages 270–287, San Diego, CA, 1990.

[121] S. R. Arridge. Photon-measurement density functions. Part 1: Analytical forms. *Applied Optics*, 34(31):7395–7409, November 1995.

APPENDICES

APPENDIX A
PROOF OF MULTIGRID MONOTONE CONVERGENCE

We begin with two lemmas which give sufficient conditions to guarantee monotone decrease of the finer grid cost functional in the two-grid algorithm. All lemmas assume that the functions $\tilde{c}^{(q)}(\cdot)$ and $\tilde{c}^{(q+1)}(\cdot)$ are continuously differentiable.

Lemma 1: Assume that the following conditions are satisfied for a resolution $q \geq 0$:

1. The fixed grid update is monotone at resolutions q and $q+1$.
2. A functional $\eta^{(q+1)} : I\!R^{N^{(q+1)}} \rightarrow I\!R$, defined by

$$\eta^{(q+1)}(x^{(q+1)}) = c^{(q+1)}(x^{(q+1)}) - c^{(q)}(x^{(q)} + I^{(q)}_{(q+1)}(x^{(q+1)} - I^{(q+1)}_{(q)}x^{(q)})) \ , \quad \text{(A.1)}$$

   has a global minimum at $x^{(q+1)} = I^{(q+1)}_{(q)}x^{(q)}$, where $x^{(q)}$ is the value resulting after the initial $\nu_1^{(q)}$ fixed grid iterations.
3. $\nu_1^{(q)} + \nu_2^{(q)} \geq 1$.

Then, the two-grid inversion algorithm of Fig. 2.2 is monotone for the functional $c^{(q)}(\cdot)$.

Proof of Lemma 1:

By the definition of monotonicity, the updated value $\tilde{x}^{(q+1)}$ of (2.9) satisfies

$$c^{(q+1)}(\tilde{x}^{(q+1)}) \leq c^{(q+1)}(I^{(q+1)}_{(q)}x^{(q)}) \ . \quad \text{(A.2)}$$

Applying the definition of $\eta^{(q+1)}(\,\cdot\,)$ and the second condition, we have

$$\eta^{(q+1)}(\tilde{x}^{(q+1)}) \ \geq \ \eta^{(q+1)}(I^{(q+1)}_{(q)}x^{(q)})$$

or equivalently

$$c^{(q+1)}(\tilde{x}^{(q+1)}) - c^{(q)}(x^{(q)} + I^{(q)}_{(q+1)}(\tilde{x}^{(q+1)} - I^{(q+1)}_{(q)}x^{(q)})) \ \geq \ c^{(q+1)}(I^{(q+1)}_{(q)}x^{(q)}) - c^{(q)}(x^{(q)}) \quad \text{(A.3)}$$

From the inequalities (A.2) and (A.3), it follows that

$$c^{(q)}(x^{(q)} + I^{(q)}_{(q+1)}(\tilde{x}^{(q+1)} - I^{(q+1)}_{(q)}x^{(q)})) \;\leq\; c^{(q)}(x^{(q)}) \; . \tag{A.4}$$

This inequality means that the coarse grid update and its subsequent coarse grid correction decreases the cost functional $c^{(q)}(\,\cdot\,)$. With the first condition, this guarantees the inequality in the definition of monotone convergence for $c^{(q)}(\,\cdot\,)$. Furthermore, by the first and fourth conditions, if $\nabla c^{(q)}(x^{(q)}) \neq 0$, the update at resolution q either before or after the coarse grid update strictly decreases $c^{(q)}(\,\cdot\,)$. Therefore, the two-grid algorithm is monotone under these assumptions.

Lemma 2. (Two-Grid Monotone Convergence)

Assume that the following conditions are satisfied for a resolution $q \geq 0$:

1. The fixed grid update is monotone at resolutions q and $q+1$.
2. $\xi^{(q+1)}(\,\cdot\,)$ is convex on $\mathbb{R}^{N^{(q+1)}}$.
3. The adjustment vector $r^{(q+1)}$ is given by (3.15).
4. $\nu_1^{(q)} + \nu_2^{(q)} \geq 1$.

Then, the two-grid inversion algorithm of Fig. 2.2 is monotone for the functional $c^{(q)}(\cdot)$.

Proof of Lemma 2:

It is enough to show that the third and second conditions of this lemma imply the second condition in Lemma 1. By condition three, we know that

$$\eta^{(q+1)}(x^{(q+1)}) = \xi^{(q+1)}(x^{(q+1)}) + vx^{(q+1)} + \text{constant} \tag{A.5}$$

for some row vector v of length $N^{(q+1)}$. In fact, we know that equation (2.15) selects the vector v so that the gradients of the coarse and fine scale cost functionals are matched, and therefore

$$\nabla\eta^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I^{(q+1)}_{(q)}x^{(q)}} = 0 \; . \tag{A.6}$$

By (A.5) we also know that $\eta^{(q+1)}(\cdot)$ is a continuously differentiable convex function. Therefore, we know that $\eta^{(q+1)}(\cdot)$ must take on its global minimum value at $x^{(q+1)} = I^{(q+1)}_{(q)}x^{(q)}$.

Proof of Multigrid Monotone Convergence Theorem:

Our proof is by induction. Consider the case when $q = Q - 2$, then we have the two-grid case, and the proof is trivial by Lemma 2.

Now consider $q < Q-2$. By induction, we assume that the Multigrid-V algorithm applied at resolution $q + 1$ is monotone for the function $c^{(q+1)}(\cdot)$. This then meets condition 1 of Lemma 2, since the multigrid algorithm serves as the coarse grid optimizer in a two-grid algorithm. Therefore, the Multigrid-V algorithm applied at resolution q is monotone for the function $c^{(q)}(\cdot)$, and the induction is complete.

APPENDIX B
COMPUTATIONAL COMPLEXITY OF MULTIGRID INVERSION

In this appendix, we compare the computational cost of the proposed multigrid inversion algorithm for ODT problems, which is described in Chapter 2, with that of the fixed-grid ICD algorithm [77]. We use the number of complex multiplications required for one iteration of the V-cycle algorithm as a measure of computational complexity.

First, let us consider the computation required for one iteration of Fixed_Grid_Update(). Here, we use the analysis from [77]. Assuming F iterations are used for the linear PDE solver, the computation of Green's functions of (2.40) and (2.41) needs $5(K+M)FN_0$ multiplications, where N_0 is the number of grid points in the PDE domain. Then, we need PN and $\frac{5}{2}PN$ multiplications to compute (2.39) and (2.44), respectively, where N is the updated image size. [1] Thus, the total computational cost for one iteration of the ICD fixed-grid update is $5(K+M)FN_0 + \frac{7}{2}PN$ multiplications.

Now, let us estimate the computation required for one iteration of MultigridV() which operates at resolutions $0, \ldots, Q-1$. For simplicity, we neglect the computational cost required for decimation and interpolation of images and the correction vector. In other words, we assume that the main computational cost at resolution q consists of the fixed-grid update on $x^{(q)}$ and the computation of $r^{(q)}$. To update $x^{(q)}$, one iteration of MultigridV() involves $\nu^{(q)} = \nu_1^{(q)} + \nu_2^{(q)}$ iterations of Fixed_Grid_Update(), which requires $[5(K+M)FN_0^{(q)} + \frac{7}{2}PN^{(q)}]\nu^{(q)}$ multiplica-

[1] In Section 2.4, we do not update the outermost region to avoid singularity problems, so N and N_0 are different in this case.

tions. Since $N_0^{(q)} = 8^{-q}N_0$ and $N^{(q)} = 8^{-q}N$ in 3-D problems, this is equal to $8^{-q} \times [5(K+M)FN_0 + \frac{7}{2}PN]\nu^{(q)}$ multiplications.

The correction vector $r^{(q)}$ is computed only once when the inversion proceeds from resolution q to $q+1$. Since $g^{(q+1)}$ is computed in the optimization for the update of $x^{(q+1)}$, the only additional computation for $r^{(q)}$ is computation of $g^{(q)}$ given by (2.17). To compute $g^{(q)}$, we first compute the Green's functions of (2.40) and (2.41) and then use them to compute Fréchet derivatives by (2.39), which requires $5(K+M)FN_0^{(q)}$ and $PN^{(q)}$ multiplications, respectively [77]. Then, $PN^{(q)}$ multiplications are required to evaluate the expression in the braces of (2.17). The resulting total complexity for computation of $r^{(q)}$ is $8^{-q} \times [5(K+M)FN_0 + 2PN]$ multiplications.

Thus, for resolutions $q = 0, \ldots, Q-2$, the total complexity of the Multigrid-V algorithm is $8^{-q} \times [\{5(K+M)FN_0 + \frac{7}{2}PN\}\nu^{(q)} + \{5(K+M)FN_0 + 2PN\}]$ multiplications. At the coarsest resolution $q = Q-1$, we do not need $r^{(Q-1)}$, so the complexity is $8^{-(Q-1)} \times \{5(K+M)FN_0 + \frac{7}{2}PN\}\nu^{(Q-1)}$ multiplications. Therefore, the total complexity for one Multigrid-V is

$$\sum_{q=0}^{Q-2} \left[8^{-q} \times \left\{ \left\{ 5(K+M)FN_0 + \frac{7}{2}PN \right\} \nu^{(q)} + \{5(K+M)FN_0 + 2PN\} \right\} \right] + 8^{-(Q-1)} \times \left\{ 5(K+M)FN_0 + \frac{7}{2}PN \right\} \nu^{(Q-1)} , \quad \text{(B.1)}$$

where K is the number of sources, M is the number of detectors, P is the number of measurements, N_0 is the PDE image size, N is the updated image size, F is the number of iterations required for the linear forward solver, and $\nu^{(q)}$ is the number of iterations of fixed grid update at resolution q.

Table 2.3 lists the estimated number of complex multiplications required for each iteration of the fixed-grid and Multigrid-V algorithms for typical values of parameters which we use in the simulations of Section 2.4.2. The values of the parameters are $K = 48$, $M = 54$, $P = 2160$, $N_0 = 65 \times 65 \times 65$, $N = 49 \times 49 \times 49$, and $F = 16$. We also provide the experimental computation time. One fixed-grid iteration took 55.5 minutes of user time on a Pentium-III 697 MHz Linux machine, and the complexity per iteration is $4.56 \sim 4.96$ times larger for the multigrid algorithm. However, one

multigrid iteration involves many coarser grid iterations, and the simulation results show that the number of iterations required for the multigrid algorithms to converge is substantially less than is required using the fixed grid algorithm.

APPENDIX C
COMPUTATIONAL COMPLEXITY OF MULTIGRID INVERSION WITH VARIABLE DATA RESOLUTION

In this appendix, we analyze the computational cost of the multigrid inversion algorithms described in Chapter 3. We use the number of multiplications/divisions (and the number of additional exponentiations in the Poisson transmission case) as a measure of computational complexity.

For simplicity, we make three assumptions. First, all the data-independent vectors and matrices, such as P, Λ, and a, are precomputed and stored. Second, the ratio M_0/M is approximately constant across resolutions, where M_0 is the average number of nonzero projections associated with each image pixel. Finally, we neglect the computational cost required for decimation and interpolation. In other words, we assume that the main computational cost at resolution q consists of the fixed-grid update on $x^{(q)}$ and the computation of $r^{(q)}$.

The ICD iteration typically has complexity of $O(M_0 N)$, where N is the number of pixels. Thus, one ICD iteration at scale q requires only 16^{-q} times the computations at the finest scale for the variable data resolution case, and 4^{-q} times the computation at the finest scale for fixed data resolution case. This is also true for the $r^{(q)}$ computation, which is computed only once when the inversion proceeds from scale q to $q+1$.

Then, in a similar manner to [91], the complexity of one MultigridV iteration is given by

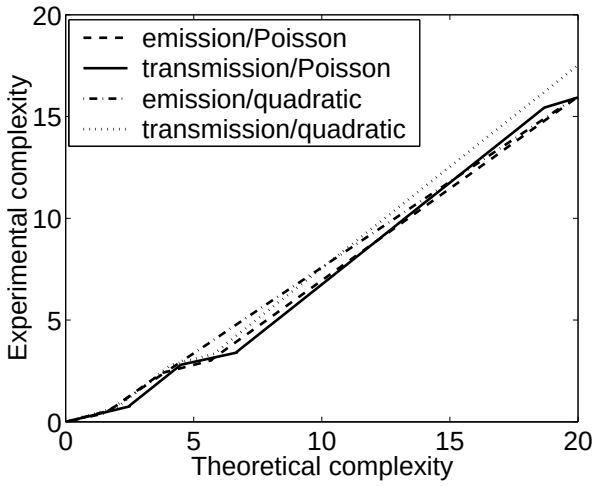
$$\sum_{q=0}^{Q-2} \left[16^{-q} \times Comp_x \times \nu^{(q)} + Comp_r \right] + 16^{-(Q-1)} \times Comp_x \times \nu^{(Q-1)} \qquad \text{(C.1)}$$

for the variable data resolution case, and

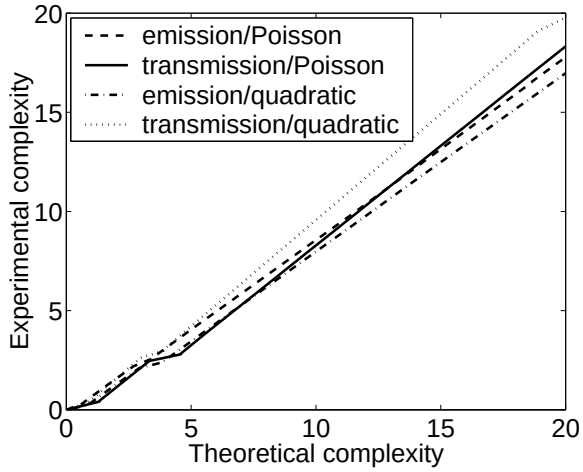
$$\sum_{q=0}^{Q-2} \left[4^{-q} \times Comp_x \times \nu^{(q)} + Comp_r \right] + 4^{-(Q-1)} \times Comp_x \times \nu^{(Q-1)} \qquad \text{(C.2)}$$

for fixed data resolution case, where $Comp_x$ is the complexity for one ICD iteration at the finest scale, $Comp_r$ is the complexity for updating r vector at the finest scale, and $\nu^{(q)} = \nu_1^{(q)} + \nu_2^{(q)}$ is the number of iterations of fixed grid update at scale q. The ratio of $Comp_r$ to $Comp_x$ is $\frac{2}{3}$ for the quadratic cases; $\frac{2}{5}$ for the Poisson emission case; and 1 for the Poisson transmission case, where we conservatively assume that the exponentiations dominate the complexity. The formulas (C.1) and (C.2) were used to scale the iteration number in Sec. 3.5.

Figure C.1 compares the theoretical complexity, computed with (C.1) and (C.2), with a measured experimental complexity in terms of the CPU time. The experimental complexity is the CPU time divided by the average CPU time of one fixed-grid ICD iteration. It was measured on a linux machine with an AMD 2.0 GHz Athlon CPU and 2 GByte memory. The experimental complexity for multigrid algorithms was consistently a little lower than the theoretical complexity. Interestingly, we found that the coarse scale ICD iterations took substantially shorter time than the theoretical complexity anticipates, which might be an effect of the better cache locality when solving the small scale problem.



(a)



(b)

Fig. C.1. Comparison between the theoretical complexity and the measure CPU time for the multigrid algorithms with (a) fixed data resolution and (b) variable data resolution

APPENDIX D
MULTIGRID INVERSION WITH VARIABLE DATA RESOLUTION FOR GAUSSIAN DATA WITH NOISE SCALING PARAMETER ESTIMATION

In this appendix, we describe a multigrid inversion method with variable data resolution, which is applicable for Gaussian noise data with automatic estimation of the noise scaling parameter. More specifically, the cost function is given by

$$c(x) = M \log ||y - f(x)||^2_\Lambda + S(x), \tag{D.1}$$

as described in Chapter 2. We have found that the cost function with the logarithm is helpful for robust convergence in nonconvex optimization arising from highly nonlinear inverse problems, such as ODT. However, in such inverse problems, the assumption (3.5) is not generally satisfied. For example, we showed that for the ODT problem, the discretization error in the forward model evaluation is not negligible compared to the measurement noise [91]. Thus, applications of the method presented in Sec. 3.2.1 to highly nonlinear inverse problems can be problematic.

In this section, we present a multigrid inversion algorithm with variable data resolution for the cost function (D.1). Basically, we apply the method presented in [91], but considering variable dimensions of y, $f(\cdot)$, and Λ.

We define cost function, with a form analogous to that of (3.4), but with quantities indexed by the scale q and appending an additional linear correction term, as

$$c^{(q)}(x^{(q)}) = M \log ||y^{(q)} - f^{(q)}(x^{(q)})||^2_{\Lambda^{(q)}} + S^{(q)}(x^{(q)}) - r^{(q)}x^{(q)} \ , \tag{D.2}$$

where $r^{(q)}$ is a row vector used to adjust the function's gradient. At the finest scale, all quantities take on their fine scale values and $r^{(q)} = 0$, so that $c^{(0)}(x^{(0)}) = c(x)$.

The forward model $f^{(q)}(\cdot)$ and the stabilizing function $S^{(q)}(\cdot)$ can be chosen in the same manner to Sec. 3.2.1. The quantity $y^{(q)}$ denotes an adjusted measurement vector at scale q.

We choose a coarse scale cost function which matches the fine cost function, as described in (3.13), by adjusting $y^{(q+1)}$ and $r^{(q+1)}$ dynamically when proceeding from scale q to $q+1$, and precomputing $\Lambda^{(q)}$. First, we make the initial error between the forward model and measurements at the coarse scale the same as the decimated fine scale error. The condition can be expressed as

$$y^{(q+1)} - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) = J_{(q)}^{(q+1)} \left[y^{(q)} - f^{(q)}(x^{(q)}) \right] \tag{D.3}$$

at the current value of $x^{(q)}$. This yields the update for $y^{(q+1)}$

$$y^{(q+1)} \leftarrow J_{(q)}^{(q+1)} y^{(q)} - \left[J_{(q)}^{(q+1)} f^{(q)}(x^{(q)}) - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) \right] \ . \tag{D.4}$$

Intuitively, the term in the bracket compensates for the forward model mismatch between resolutions. In a special case when $J_{(q)}^{(q+1)} = I_{(q)}^{(q+1)}$, the measurement vector update (3.26) becomes

$$y^{(q+1)} \leftarrow I_{(q)}^{(q+1)} y^{(q)} - \left[I_{(q)}^{(q+1)} f^{(q)}(x^{(q)}) - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) \right] \ , \tag{D.5}$$

which is exactly same as the way how the full approximation scheme (FAS) [27, 28] compensates for equation mismatch between scales.

Second, we choose the coarse scale weight matrix $\Lambda^{(q)}$

$$\Lambda^{(q+1)} \triangleq \left[J_{(q+1)}^{(q)} \right]^T \Lambda^{(q)} J_{(q+1)}^{(q)} \ . \tag{D.6}$$

Note that $\Lambda^{(q+1)}$ is independent of the image, and thus can be precomputed.

Finally, we use the gradient match condition (3.14). More specifically, the gradient adjustment factor $r^{(q+1)}$ is computed by (3.16), where $g^{(q)}$ and $g^{(q+1)}$ are given by

$$\begin{aligned} g^{(q)} = \ & -\frac{2M}{||y^{(q)} - f^{(q)}(x^{(q)})||^2_{\Lambda^{(q)}}} \left(y^{(q)} - f^{(q)}(x^{(q)}) \right)^T \Lambda^{(q)} A^{(q)} \\ & + \nabla S^{(q)}(x^{(q)}) \end{aligned} \tag{D.7}$$

$$g^{(q+1)} = -\frac{2M}{||y^{(q+1)} - f^{(q+1)}(x^{(q+1)})||^2_{\Lambda^{(q+1)}}} \left(y^{(q+1)} - f^{(q+1)}(x^{(q+1)})\right)^T \Lambda^{(q+1)} A^{(q+1)} + \nabla S^{(q+1)}(I^{(q+1)}_{(q)} x^{(q)}), \tag{D.8}$$

where T is the transpose operator, and $A^{(q)}$ denotes the gradient of the forward model or Fréchet derivative given by $A^{(q+1)} = \nabla f^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I^{(q+1)}_{(q)}x^{(q)}}$ and $A^{(q)} = \nabla f^{(q)}(x^{(q)})$.

The multigrid recursions with variable data resolution for Gaussian data with noise scaling parameter estimation is summarized in the pseudocodes in Fig. D.1. Note that the main difference of this algorithm from that in Fig. 3.1 is that coarse scale measurement vector is also dynamically adjusted to compensate for the forward model mismatch.

```
main( ) {
    Initialize x^(0) with a background estimate
    r^(0) ← 0
    y^(0) ← y
    For q = 1, 2, ..., Q − 1,        compute Λ^(q) using (D.6)
    Choose number of fixed grid iterations ν_1^(0), ..., ν_1^(Q−1) and ν_2^(0), ..., ν_2^(Q−1)
    Repeat until converged:
        x^(0) ← MultigridV(q, x^(0), c^(0)( · ; y^(0), r^(0)))
}
```

(a)

```
x^(q) ← MultigridV(q, x^(q), y^(q), r^(q)) {
    Repeat ν_1^(q) times
        x^(q) ← Fixed_Grid_Update(x^(q), c^(q)( · ; y^(q), r^(q))) //Fine grid update
    If q = Q − 1, return x^(q) //If coarsest scale, return result
    x^(q+1) ← I_(q)^(q+1) x^(q) //Decimation
    Compute y^(q+1) using (D.4)
    Compute r^(q+1) using (3.16), (D.7), and (D.8)
    x^(q+1) ← MultigridV(q + 1, x^(q+1), y^(q+1), r^(q+1)) //Coarse grid update
    x^(q) ← x^(q) + I_(q+1)^(q) (x^(q+1) − I_(q)^(q+1) x^(q)) //Coarse grid correction
    Repeat ν_2^(q) times
        x^(q) ← Fixed_Grid_Update(x^(q), c^(q)( · ; y^(q), r^(q))) //Fine grid update
    Return x^(q) //Return result
}
```

(b)

Fig. D.1. Pseudo-code specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion for Gaussian data with unknown noise scaling parameter estimation

VITA

VITA

Seungseok Oh received the B.S. and M.S. degrees in Electrical Engineering from Seoul National University, Seoul, Korea, in 1997 and 1999, respectively. In 1999-2000, he was with the Hanaro Telocom, Inc. as a network engineer. He is currently working toward the Ph.D. degree in the School of Electrical and Computer Engineering, Purdue University, West Lafayette. His current research interests include image processing, inverse problems, medical imaging, multimedia systems, and biomedical optics.