NONLINEAR MULTIGRID OPTIMIZATION FOR BAYESIAN DIFFUSION TOMOGRAPHY

Jong Chul Ye, Charles A. Bouman, Rick P. Millane and Kevin J. Webb

Purdue University
West Lafayette, IN 47907-1285
{jong,bouman,webb}@ ecn.purdue.edu

ABSTRACT

Optical diffusion tomography attempts to reconstruct an object cross section from measurements of scattered and attenuated light. While Bayesian approaches are well suited to this difficult nonlinear inverse problem, the resulting optimization problem is very computationally expensive. In this paper, we propose a nonlinear multigrid technique for computing the maximum a posteriori (MAP) reconstruction in the optical diffusion tomography problem. The multigrid approach improves reconstruction quality by avoiding local minimum. In addition, it dramatically reduces computation. Each iteration of the algorithm alternates a Born approximation step with a single cycle of a nonlinear multigrid algorithm.

1. INTRODUCTION

The objective of optical diffusion imaging is to reconstruct the cross section of a highly scattering media such as tissue based on measurements of the scattered and attenuated optical energy. This technique presents significantly lower health risk as compared to X-ray imaging techniques and its potential has been successfully demonstrated in biomedical applications [1]. However, the relationship between the unknown scattering coefficients and the optical measurements is highly nonlinear and described by a partial differential equation; so reconstruction poses a very challenging nonlinear inverse problem.

Recently, Bayesian approaches based on iterative coordinate descent (ICD) optimization have been investigated [2, 3]. In particular, we have proposed an ICD/Born method which provides high quality reconstructions and is computationally efficient when compared to the conventional iterative Born approximation methods. However, the computational complexity of ICD/Born is still prohibitive for large three dimensional problems. Perhaps more importantly, local

optimization methods such as ICD/Born can become trapped in local minimum and result in suboptimal solutions.

In this paper, we apply multigrid algorithms to nonlinear optical diffusion imaging. Multigrid techniques have a number of important advantages over fixed grid methods for our type of problem [4]. First, the convergence of multigrid algorithms is independent of the size of the image, reducing overall computation. Second, the global nature of the optimization tends to more robustly achieve the global minimum, resulting in improved reconstruction quality. In our multigrid algorithm, each iteration alternates a Born approximation step with a single cycle of a nonlinear multigrid algorithm [5]. At each resolution of the multigrid algorithm, we use the ICD optimization method [6] (also referred to as Gauss-Seidel in multigrid applications).

We present simulation results for a 2-D reconstruction of size 129×129. The results show that the multigrid optimization reduces computation by a factor of 20-30 in our simulations while achieving substantially better reconstruction quality.

2. BAYESIAN OPTICAL DIFFUSION IMAGING

Figure 1 illustrates the proposed measurement geometry. The region to be imaged is denoted by Ω and is surrounded by K point sources at positions $s_k \in \Omega$, and M detectors at positions $d_m \in \Omega$. The following partial differential equation describes the measurement $\phi_k(r)$ at position r due to a source of light at position $s_k \in \Omega$ [1, 3].

$$\nabla \cdot D(r)\nabla\phi_k(r) + (-\mu_a(r) + j\omega/c)\phi_k(r) = -\delta(r - s_k),$$

where c is the speed of light in the medium, $D(r)$ is the diffusion constant given by $D(r) = 1/3(\mu_a(r) + \mu'_s(r))$, $\mu_a(r)$ is the absorption coefficient, and $\mu'_s(r)$ is the reduced scattering coefficient. Here $\phi_k(r)$ denotes the complex envelope of a modulated optical signal [1, 3]. Then, the objective of the optical imaging problem is to

determine the values of $\mu_a(r)$ from the measured values of $\phi_k(d_m)$. This is a difficult nonlinear ill-posed inverse problem because of the nonlinear coupling between the coefficients and the photon flux in the diffusion equation.

The Bayesian framework for maximum *a posteriori* (MAP) estimation of absorption image $\mathbf{x} \in \mathbb{R}^N$ from the measurement vector $\mathbf{y} \in \mathbb{C}^P$ is given by

$$\hat{\mathbf{x}}_{MAP} \quad = \quad \arg\max_{\mathbf{x}} \{ \ \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) \ \}, \quad (1)$$

where $p(\mathbf{y}|\mathbf{x})$ is the data likelihood, and $p(\mathbf{x})$ is the prior density for the image. Under appropriate conditions, we showed that the data likelihood can be represented as [3]

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

where α is a parameter which determines noise variance, Λ denotes a diagonal matrix due to "shot noise" [3], and the complex vector valued function $\mathbf{f}(\mathbf{x})$ represents the "exact" value of the photon flux for the assumed value of the absorption coefficient $\mathbf{x}$. We use the generalized Gaussian Markov random field (GGMRF) prior model of [7] since this prior model enforces smoothness in the solution while preserving sharp edge transitions.

In this paper, we will adaptively estimate α during the reconstruction procedure. In practice, we will see that adaptive estimation of α allows for smoother convergence to the minimum of our cost function. This is particularly important because $\mathbf{f}(\mathbf{x})$ is highly nonlinear so the computation of the MAP estimate can become trapped in local minima. Initially, the estimated value of α will be large when $\mathbf{x}$ is far from its true value. In this case, the strong prior term restricts the solution to be smoother. As the optimization proceeds, the value of α decreases making the data term more important and consequently reducing the relative importance of the regularization term.

If we consider α unknown, the optimization problem (1) can be re-written as

$$\min_{\mathbf{x} \geq \mathbf{0}} \min_{\alpha} \left\{ \frac{||\mathbf{y} - \mathbf{f}(\mathbf{x})||^2_\Lambda}{\alpha} + P \log \alpha + \frac{\sum b_{i-j}|x_i - x_j|^p}{p\sigma^p} \right\},$$

which can be solved by alternately minimizing with respect to α, and then $\mathbf{x}$ using the following two equations.

$$\hat{\alpha} = \frac{1}{P}||\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}})||^2_\Lambda \qquad (3)$$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \geq \mathbf{0}} \left\{ \frac{||\mathbf{y} - \mathbf{f}(\mathbf{x})||^2_\Lambda}{\hat{\alpha}} + \frac{\sum b_{i-j}|x_i - x_j|^p}{p\sigma^p} \right\}(4)$$

3. NONLINEAR MULTIGRID INVERSION

Equation (3) is a straight-forward computation, but (4) is a computationally expensive optimization problem especially for large images $\mathbf{x}$. To circumvent this problem, we will employ multigrid optimization algorithms to efficiently compute (4). At the beginning of each multigrid cycle the nonlinear functional $\mathbf{f}(\mathbf{x})$ is first linearized using a Taylor series expansion (or Born approximation)

$$||\mathbf{y} - \mathbf{f}(\mathbf{x})||^2_\Lambda \quad \simeq \quad ||\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}) - \mathbf{f}'(\hat{\mathbf{x}})\Delta\mathbf{x}||^2_\Lambda \quad (5)$$

where $\hat{\mathbf{x}}$ is an estimate of the unknown $\mathbf{x}$, $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$, and $\mathbf{f}'(\hat{\mathbf{x}})$ represents the Fréchet derivative of $\mathbf{f}(\cdot)$ at $\hat{\mathbf{x}}$ [3]. This presents an approximate cost function

$$c(\mathbf{x}) \quad = \quad ||\mathbf{z} - \mathbf{A}\mathbf{x}||^2_\mathbf{C} + \frac{\sum b_{i-j}|x_i - x_j|^p}{p\sigma^p}, \quad (6)$$

where

$$\mathbf{z} = \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}) + \mathbf{f}'(\hat{\mathbf{x}})\hat{\mathbf{x}}, \quad \mathbf{A} = \mathbf{f}'(\hat{\mathbf{x}}), \quad \mathbf{C} = \frac{\Lambda}{\alpha} \quad (7)$$

The next section describes how the approximate cost function of (6) can be efficiently minimized using a multigrid algorithm. In particular, we will investigate the use of two different multigrid methods known as V-cycle and full multigrid [4].

Our overall strategy for the optimization is illustrated in the Figure 2. Each iteration of our algorithm starts with an update of α using (3) followed by a new linearization (6). This results in a non-linear optimization problem that we then solve using either V-cycle or full multigrid. This sequence is repeated until the desired level of convergence is reached.

3.1. Two Grid Algorithm

Let $\mathbf{x}^{(0)}$ denote the finest grid absorption image, i.e., $\mathbf{x} = \mathbf{x}^{(0)}$, and let $\mathbf{x}^{(1)}$ be the next coarser grid absorption image which is obtained by a 2:1 decrease in grid resolution. We may compute $\mathbf{x}^{(1)}$ from $\mathbf{x}^{(0)}$ via the decimation operation $\mathbf{x}^{(1)} = \mathbb{I}^{(1)}_{(0)}\mathbf{x}^{(0)}$. Here $\mathbb{I}^{(1)}_{(0)}$ is an $\frac{N}{4} \times N$ matrix which locally averages points in $\mathbf{x}^{(0)}$ to form $\mathbf{x}^{(1)}$. For the two dimensional image, we choose $\mathbb{I}^{(1)}_{(0)}$ to be the separable extension of the one dimensional decimation matrix

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}. \quad (8)$$

For most common choices of matrices used on a two dimensional grid, the corresponding interpolation matrix $\mathbb{I}^{(0)}_{(1)}$ is defined by $\mathbb{I}^{(1)}_{(0)} = \frac{1}{4}(\mathbb{I}^{(0)}_{(1)})^T$.

Multigrid optimization can be best understood by first considering the two-grid algorithm. Assume that we have an initial solution $\hat{\mathbf{x}}^{(0)}$ which approximately minimizes the cost functional

$$c^{(0)}(\mathbf{x}^{(0)}) = ||\mathbf{z} - \mathbf{A}\mathbf{x}^{(0)}||^2_\mathbf{C} + \frac{\sum b_{i-j}\left|x_i^{(0)} - x_j^{(0)}\right|^p}{p\sigma^p} \quad . \quad (9)$$

Our objective is to then improve our result by incorporating $\mathbf{x}^{(1)}$, an improved solution computed at a coarse grid. We may do this by using the update formula

$$\mathbf{x}^{(0)} \quad \leftarrow \quad \hat{\mathbf{x}}^{(0)} + \mathbb{I}^{(0)}_{(1)}(\mathbf{x}^{(1)} - \mathbb{I}^{(1)}_{(0)}\hat{\mathbf{x}}^{(0)}) \quad . \quad (10)$$

In order to implement this coarse grid correction, we must first define a corresponding cost function, $c^{(1)}(\mathbf{x}^{(1)})$, for the coarse grid solution $\mathbf{x}^{(1)}$. Then, $c^{(1)}(\mathbf{x}^{(1)})$ consists of two parts: a quadratic term and a prior term. Referring to (10), the quadratic term of (9) can be expressed as

$$||\mathbf{z} - \mathbf{A}\mathbf{x}^{(0)}||^2_\mathbf{C} \quad = \quad ||\mathbf{z}^{(1)} - \mathbf{A}^{(1)}\mathbf{x}^{(1)}||^2_\mathbf{C} \quad (11)$$

where $\mathbf{z}^{(1)} = \mathbf{z} - \mathbf{A}\left(1 - \mathbb{I}^{(0)}_{(1)}\mathbb{I}^{(1)}_{(0)}\right)\hat{\mathbf{x}}^{(0)}$ and $\mathbf{A}^{(1)} = \mathbf{A}\mathbb{I}^{(0)}_{(1)}$.

For the coarse grid prior term, we assume that the derivative of x is locally smooth [5]. In this case, we use the approximation

$$\sum b_{i-j}\left|x_i^{(0)} - x_j^{(0)}\right|^p \simeq 4\sum b_{i-j}\left|\frac{x_i^{(1)} - x_j^{(1)}}{2}\right|^p (12)$$

Note that the factor of 4 is chosen to account for the reduced number of terms in the sum, and the smoothness assumption makes $x_i^{(0)} - x_j^{(0)} \simeq (x_i^{(1)} - x_j^{(1)})/2$. Based on (11) and (12), we define the coarse grid cost function as

$$\begin{aligned} c^{(1)}(\mathbf{x}^{(1)}) &= &||\mathbf{z}^{(1)} - \mathbf{A}^{(1)}\mathbf{x}^{(1)}||^2_\mathbf{C} \\ &+ &\frac{4}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|\frac{x_i^{(1)} - x_j^{(1)}}{2}\right|^p .(13) \end{aligned}$$

However, minimization of $c^{(1)}(\mathbf{x}^{(1)})$ is still problematic because discretization errors for the prior term will cause the solution to be slightly different than would otherwise result from the exact fine grid optimization of (9). To correct for this error, we will solve the optimization problem

$$\min_{\mathbf{x}^{(1)}\geq\mathbf{0}}\left\{c^{(1)}(\mathbf{x}^{(1)}) - (\mathbf{r}^{(1)})^T\mathbf{x}^{(1)}\right\} \quad , \quad (14)$$

where the vector $\mathbf{r}^{(1)}$ is chosen as a correction term so that the exact fine grid solution is a fixed point to the new coarse grid problem. That is

$$\nabla c^{(1)}\left(\mathbb{I}^{(1)}_{(0)}\hat{\mathbf{x}}^{(0)}\right) - \mathbf{r}^{(1)}\begin{cases}= \mathbf{0} \text{, if } \hat{\mathbf{x}}^{(0)} \text{ is the solution} \\ \neq \mathbf{0} \text{, otherwise}\end{cases} \quad (15)$$

Furthermore, we want $\mathbf{r}^{(1)}$ to approach zero when the approximation error due to the discretization of the prior term is small. We can show that the following expression for $\mathbf{r}^{(1)}$ satisfies these two conditions.

$$\mathbf{r}^{(1)} \quad = \quad \nabla c^{(1)}(\mathbb{I}^{(1)}_{(0)}\hat{\mathbf{x}}^{(0)}) - 4\mathbb{I}^{(1)}_{(0)}\nabla c^{(0)}(\hat{\mathbf{x}}^{(0)}). \quad (16)$$

Then, the coarse grid problem can be represented as

$$\begin{aligned} \min_{\mathbf{x}^{(1)}\geq\mathbf{0}}&\left\{||\mathbf{z}^{(1)} - \mathbf{A}^{(1)}\mathbf{x}^{(1)}||^2_\mathbf{C} - (\mathbf{r}^{(1)})^T\mathbf{x}^{(1)}\right. \\ &\left.+\frac{4}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|\frac{x_i^{(1)} - x_j^{(1)}}{2}\right|^p\right\} . \quad (17) \end{aligned}$$

Multigrid optimization is implemented by recursively applying the two-grid optimization. In particular, we will use the two recursions known as V-cycle and full multigrid [4]. The pseudo-code recursions for V-cycle and full multigrid are shown in Tables 1 and 2.

At each resolution or grid, we use the ICD optimization method [6, 2, 3] to compute an approximate solution to the optimization. The ICD method is essentially a fast implementation of the Gauss-Seidel method for tomography applications. In fact, Gauss-Seidel optimization is usually the preferred method for use in multigrid optimization because it has fast convergence at high spatial frequencies. This is also true for the case of tomographic reconstruction, so the ICD method is well suited to our problem.

4. SIMULATION RESULTS

We used an 8 point neighborhood model with $b_{i-j} = (2\sqrt{2} + 4)^{-1}$ for nearest neighbors and $b_{i-j} = (4\sqrt{2} + 4)^{-1}$ for diagonal neighbors. A fixed value of 1.1 is used for p in all the reconstructions. A random update ordering scheme was used for the ICD relaxation. The number of resolutions for the multigrid inversion algorithms is determined by decimating the full resolution image at scale 0 with size of 129×129 pixels until we obtain 17×17 pixels. Two iterations of optimization were applied at each resolution of the multigrid algorithm.

Figure 3(a) shows a phantom used for one of the numerical experiments. The values of the absorption coefficient for each sub-domain are given in Figure 3(b). The modulation frequency is 200 MHz. Figure 3(c)(d) illustrate the magnitude and the phase information of the scattering data. The reconstructions by the fixed grid ICD/Born after 1000 iterations (8,923 sec of CPU time on a Sun Ultra Sparc 30 machine), and by the V-cycle algorithm after 20 cycles (352 sec of CPU time) are compared in Figure 4. Notice that the multigrid reconstruction is much more accurate than the fixed resolution result. Furthermore, the multigrid algorithm reduces computation by a factor of 20-30. Figure 5(a)(b)

show the convergence of the log posterior probability and the normalized root mean square (NRMSE) values for the multigrid inversion algorithm and the fixed grid ICD/Born algorithm, as a function of CPU time. The multigrid inversion algorithm converges at a very early stage of the iterations. Figure 6 illustrates the results of 10 cycles of full multigrid (220-240 sec of CPU time) on three test phantoms each with 129×129 resolution. In each case, accurate reconstructions were obtained.

5. CONCLUSIONS

In this work, we derived a fast inversion algorithm for the Bayesian diffusion imaging using the nonlinear multigrid optimization technique. Simulation results show that the multigrid algorithms dramatically reduce the computational burden as well as improve the reconstruction quality.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] B. W. Pogue, M. S. Patterson, H. Jiang, and K. D. Paulsen, "Initial assessment of a simple system for frequency domain diffuse optical tomography," *Phys. Med. Biol.*, vol. 40, pp. 1709–1729, 1995.

[2] H. Carfantan, A. Mohammad-Djafari, and J. Idier, "A single site update algorithm for nonlinear diffraction tomography," *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, vol. 4, April 21-24 1997, Munich, Germany, pp. 2837–2840.

[3] 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," *accepted for publication, J. Opt. Soc. Am. A*, 1999.

[4] W. Briggs, *A Multigrid Tutorial*. Philadelphia: Society for Industrial and Applied Mathematics, 1987.

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

[6] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[7] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.
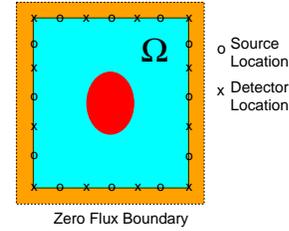
Figure 1: Simulation geometry with the locations of sources and detectors for inversion of synthetic data.
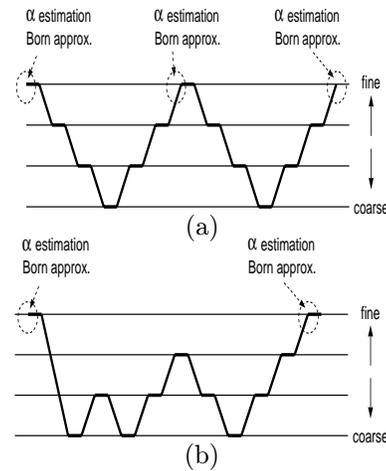

Figure 2: (a) V-cycle inversion algorithm, and (b) full multigrid inversion algorithm.
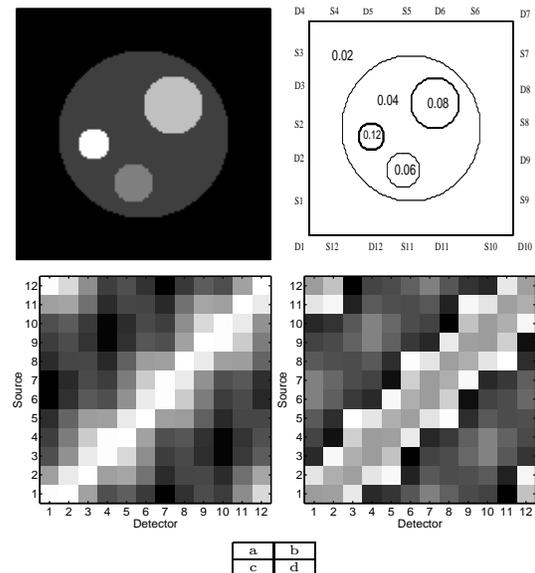

Figure 3: (a) Gray scale of absorption coefficient, and (b) contour plot with units of cm^{-1}. Gray-scale view of (c) log magnitude and (d) phase of measurements for each source and detector pairs.
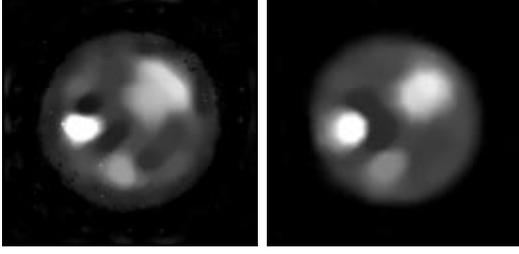
Figure 4: Reconstructions by (a) the fixed resolution ICD/Born algorithm, and (b) the multigrid V-cycle inversion algorithm.
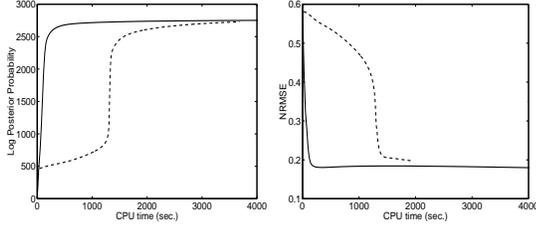

Figure 5: (a) Log posterior probability and (b) NRMSE as a function of CPU time for fixed grid ICD/Born algorithm $(\cdots)$ and multigrid V-cycle algorithm $(-)$.
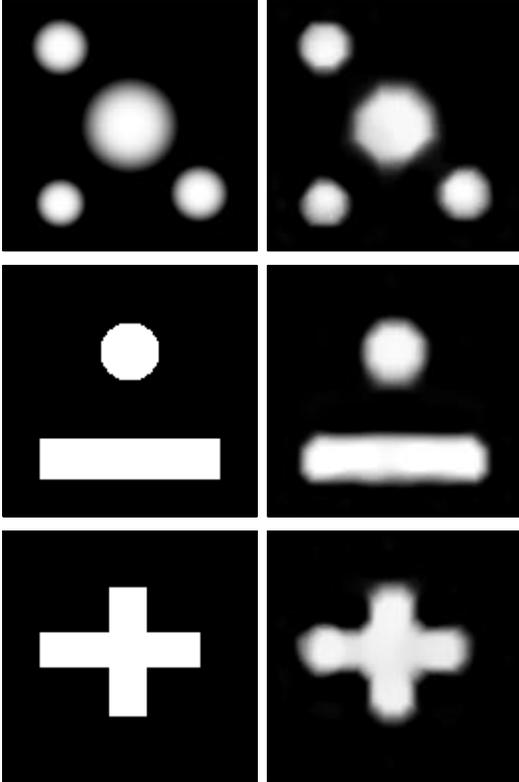

Figure 6: Reconstruction results for μ_a by the full multigrid inversion algorithm. Left column: original absorption images; right column: reconstructed absorption images after 10 cycles of full multigrid algorithm.

MultigridV$(\mathbf{x}, \mathbf{r}, \mathbf{z}, k)$ {

1. Use the initial condition, $\mathbf{x}$, to approximately solve the optimization using ν_1 scans of the ICD algorithm

$$\mathbf{x} \leftarrow \arg\min_{\mathbf{x} \geq \mathbf{0}} \left\{ ||\mathbf{z}^{(k)} - \mathbf{A}^{(k)}\mathbf{x}||_{\mathbf{C}} - \mathbf{r}^T\mathbf{x} + \frac{4^k}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| \frac{x_i^{(k)} - x_j^{(k)}}{2^k} \right|^p \right\}$$

2. If k is the coarsest desired grid, return.
3. Compute the following:

$$\hat{\mathbf{x}} = \mathbf{x}_{int} = \mathbb{I}_{(k)}^{(k+1)}\mathbf{x}$$
$$\hat{\mathbf{z}} = \mathbf{z} + \mathbf{A}^{(k)}\left(\mathbb{I}_{(k+1)}^{(k)}\mathbb{I}_{(k)}^{(k+1)}\mathbf{x}^{(k)} - \mathbf{x}^{(k)}\right)$$
$$\hat{\mathbf{r}} = \nabla c^{(k+1)}(\mathbf{x}_{int}) - 4\mathbb{I}_{(k)}^{(k+1)}\nabla c^{(k)}(\mathbf{x})$$

4. MultigridV$(\hat{\mathbf{x}}, \hat{\mathbf{r}}, \hat{\mathbf{z}}, k+1)$.
5. Perform coarse grid correction

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbb{I}_{(k+1)}^{(k)}(\hat{\mathbf{x}} - \mathbf{x}_{int})$$

6. Use the initial condition, $\mathbf{x}$, to approximately solve the optimization using ν_2 scans of the ICD algorithm

$$\mathbf{x} \leftarrow \arg\min_{\mathbf{x} \geq \mathbf{0}} \left\{ ||\mathbf{z}^{(k)} - \mathbf{A}^{(k)}\mathbf{x}||_{\mathbf{C}} - \mathbf{r}^T\mathbf{x} + \frac{4^k}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| \frac{x_i^{(k)} - x_j^{(k)}}{2^k} \right|^p \right\}$$

7. Return. }

Table 1: Pseudo-code specification for the multigrid V-cycle inversion algorithm

FMG$(\mathbf{x}, \mathbf{r}, \mathbf{z}, k)$ {

1. If k is the coarsest grid, go to 5.
2. Compute the following:

$$\hat{\mathbf{x}} = \mathbf{x}_{int} = \mathbb{I}_{(k)}^{(k+1)}\mathbf{x}$$
$$\hat{\mathbf{r}} = \nabla c^{(k+1)}(\mathbf{x}_{int}) - 4\mathbb{I}_{(k)}^{(k+1)}\nabla c^{(k)}(\mathbf{x})$$

3. Do FMG$(\hat{\mathbf{x}}, \hat{\mathbf{r}}, \mathbf{z}, k+1)$
4. Perform coarse grid correction

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbb{I}_{(k+1)}^{(k)}(\hat{\mathbf{x}} - \mathbf{x}_{int})$$

5. Do MultigridV$(\mathbf{x}, \mathbf{r}, \mathbf{z}, k)$.
6. Return. }

Table 2: Pseudo-code specification for the full multigrid inversion algorithm