

# Multigrid Bayesian Methods for Optical Diffusion Tomography

R.P. Millane<sup>+</sup>, J.C. Ye<sup>\*</sup>, C.A. Bouman<sup>†</sup> and K.J. Webb<sup>†</sup>

<sup>+</sup>Computational Science and Engineering Program and

<sup>†</sup>School of Electrical and Computer Engineering

Purdue University, West Lafayette, Indiana 47907, USA.

<sup>\*</sup>Department of Electrical and Computer Engineering

University of Illinois, Urbana, Illinois 61801, USA.

## Abstract

Optical diffusion imaging is a new imaging modality that promises great potential in applications such as medical imaging, environmental sensing and nondestructive testing. It presents a difficult nonlinear image reconstruction problem however. An inversion algorithm is formulated in a Bayesian framework, and an efficient optimization technique that uses iterative coordinate descent is presented. A general multigrid optimization technique for nonlinear image reconstruction problems is developed and applied to the optical diffusion imaging problem. Numerical results show that this approach improves the quality of reconstructions and dramatically decreases computation times.

Keywords: Optical diffusion imaging, Bayesian image reconstruction, generalized Gaussian Markov random field, iterative coordinate descent, multigrid

## 1. Introduction

Optical diffusion imaging is a new technique that has significant potential, particularly for soft-tissue imaging [1,2]. There are also other potential applications, including imaging through smoke and dust, and nondestructive evaluation of materials such as polymer composites. 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 equation is sufficiently accurate and provides a computationally tractable forward model (hence the name “diffusion imaging”). However, the inverse problem of reconstructing the absorption and/or scattering coefficients from measurements of the scattered light is highly nonlinear because of the nonlinear coupling between the coefficients and the photon flux in the diffusion equation. Most approaches have used the Born or distorted Born iterative method (DBIM) (e.g. Ref. 2). However, these approaches involve approximations that can lead to slow convergence and poor images [3,4]. We describe here a number of advances we have made in reconstruction algorithms for optical diffusion imaging using Bayesian regularization and multigrid optimization that improve speed and the quality of reconstructions.

## 2. Forward Problem and Bayesian Framework

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 [1]. In frequency domain

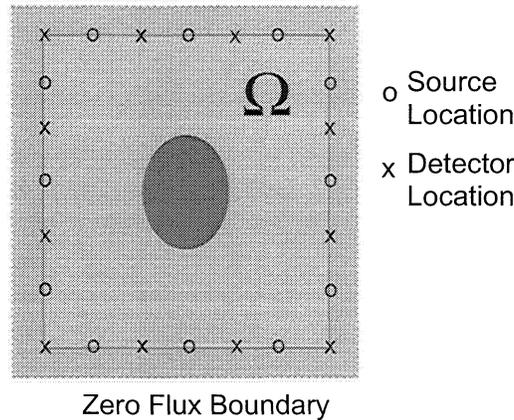


Figure 1: The region  $\Omega$  to be imaged, and the source and detector boundary.

optical diffusion imaging [2], the light source is amplitude modulated at angular frequency  $\omega$ , and the complex modulation envelope is measured at the detectors. The frequency domain method allows low noise narrow-band heterodyne detection. The complex amplitude  $\phi_k(r)$  of the modulation envelope of the optical flux density due to a point source at  $s_k$  satisfies the frequency domain diffusion equation

$$\nabla \cdot [D(r)\nabla\phi_k(r)] + (-\mu_a(r) + j\omega/c)\phi_k(r) = -\delta(r - s_k), \quad (1)$$

where  $r$  is position,  $c$  is the speed of light in the medium,  $D(r)$  is the diffusion constant given by  $D(r) = 1/3(\mu_a(r) + \mu'_s(r))$ , where  $\mu_a(r)$  is the absorption coefficient, and  $\mu'_s(r)$  is the reduced scattering coefficient. We consider a two-dimensional region to be imaged that is denoted by  $\Omega$  and is surrounded by  $K$  point sources distributed on the boundary of  $\Omega$  at positions  $s_k$ , and  $M$  detectors at positions  $d_m$  (Figure 1). Although the methodology we describe can in principle be applied in the general case of unknown absorption and scattering coefficients, for simplicity we focus on the estimation of the absorption coefficient under the assumption that the scattering coefficient is known. The reduced scattering coefficient  $\mu'_s(r)$  is then assumed to be known for all points in  $\Omega$ , but the absorption coefficient  $\mu_a(r)$  in  $\Omega$  is unknown. The domain  $\Omega$  is discretized into  $N$  pixels, where the position of the  $i$ th pixel is denoted by  $r_i$ . The set of unknown absorption coefficients is then denoted by the vector  $\mathbf{x}$  where  $\mathbf{x} = [\mu_a(r_1), \dots, \mu_a(r_N)]^T$ . Using this notation, we express the forward model as a vector valued function  $\mathbf{f}(\mathbf{x})$ , that takes on the values

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &= [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_P(\mathbf{x})]^T \\ &= [\phi_1(d_1), \phi_1(d_2) \dots, \phi_1(d_M), \phi_2(d_1), \dots, \phi_K(d_M)]^T, \end{aligned} \quad (2)$$

where the number of measurements is  $P = KM$ . The measurements of the complex envelope  $\phi_k(d_m)$  for source  $k$  and detector  $m$  are denoted by  $y_{km}$ , and are collected together into the vector  $\mathbf{y}$  of length  $P$ ,

$$\mathbf{y} = [y_{11}, y_{12}, \dots, y_{KM}]^T. \quad (3)$$

The image reconstruction problem is then to determine  $\mathbf{x}$  from  $\mathbf{y}$ . In a Bayesian framework we seek the maximum *a posteriori* (MAP) estimate of  $\mathbf{x}$ ,  $\hat{\mathbf{x}}_{MAP}$ , given by

$$\hat{\mathbf{x}}_{MAP} = \arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg \max_{\mathbf{x}} \{ \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) \}, \quad (4)$$

where  $p(\mathbf{x}|\mathbf{y})$  is the posterior density of the image given the data,  $p(\mathbf{y}|\mathbf{x})$  is the conditional probability density of  $\mathbf{y}$  given  $\mathbf{x}$ , and  $p(\mathbf{x})$  is the prior density for the image.

Using a measurement model of shot noise for the photon flux detected, we have shown [5] that the data likelihood is given by

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(\pi\alpha)^P |\Lambda|^{-1}} \exp \left[ -\frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\Lambda}^2}{\alpha} \right], \quad (5)$$

where the parameter  $\alpha$  is related to the noise variance,  $\Lambda$  is a diagonal matrix,  $\|\mathbf{z}\|_{\Lambda}^2 = \mathbf{z}^H \Lambda \mathbf{z}$ , and  $H$  denotes the Hermitian transpose. As a prior model for the image we use the generalized Gaussian Markov random field (GGMRF) model [6], since it enforces smoothness while preserving edges. For the GGMRF model, the density function for  $\mathbf{x}$  is given by

$$p(\mathbf{x}) = \frac{1}{\sigma^N z(p)} \exp \left[ -\frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \right], \quad 1 \leq p \leq 2, \quad (6)$$

where  $\mathcal{N}$  denotes all sets of neighbouring groups of pixels. Furthermore, since the absorption must be positive we also impose the constraint  $x_i \geq 0$ .

### 3. ICD/Born Optimization Technique

Referring to (4), (5) and (6), applying a positivity constraint, and (manually) setting the noise parameter  $\alpha$ , the MAP estimate for  $\mathbf{x}$  is given by

$$\hat{\mathbf{x}}_{MAP} = \arg \min_{\mathbf{x} \geq 0} \left[ \frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\Lambda}^2}{\alpha} + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \right]. \quad (7)$$

Computation of the MAP reconstruction involves performing the optimization (7). We use the ICD algorithm since it has been shown to work well with non-Gaussian prior models and is easily implemented with a positivity constraint [7]. The ICD algorithm is implemented by sequentially updating each pixel of the image. After every pixel has been updated, the procedure is repeated starting from the first pixel again. We refer to a single update of every pixel in the image as a ‘‘scan’’. The ICD algorithm therefore consists of a number of scans until some convergence criterion is satisfied. Each scan consists of  $N$  pixel updates. Each pixel update is chosen to minimize the MAP cost function, so that the update  $\hat{x}_i$  of the absorption value of the  $i$ -th pixel is given by

$$\hat{x}_i = \arg \min_{\tilde{x}_i \geq 0} \left[ \frac{\|\mathbf{y} - \mathbf{f}(\tilde{\mathbf{x}}_i)\|_{\Lambda}^2}{\alpha} + \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_i} b_{i-j} |\tilde{x}_i - x_j|^p \right], \quad (8)$$

where  $\tilde{\mathbf{x}}_i = [x_1, x_2, \dots, x_{i-1}, \tilde{x}_i, x_{i+1}, \dots, x_N]^T$  and  $\mathcal{N}_i$  is the set of pixels neighboring pixel  $i$ . Note that  $x_i$  is replaced by  $\hat{x}_i$  before the next pixel update. A direct approach [8] to the update of equation (8) is computationally expensive due to the highly nonlinear nature of the forward model  $\mathbf{f}(\mathbf{x})$ . Furthermore, each evaluation of the function  $\mathbf{f}(\mathbf{x})$  requires the solution of the full partial differential equation (1) for each source.

The computational cost is reduced by using the Born approximation at the beginning of each scan, and we call this approach the ICD/Born algorithm [5]. The integer  $n$  indexes the scans of the algorithm, and  $\mathbf{x}^n$  denotes the image after the  $n$ th scan. At the beginning of the  $(n + 1)$ th scan, the approximation

$$\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\Lambda}^2 \simeq \|\mathbf{y} - \mathbf{f}(\mathbf{x}^n) - \mathbf{f}'(\mathbf{x}^n)\Delta\mathbf{x}\|_{\Lambda}^2 \quad (9)$$

is used, where  $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^n$  and  $\mathbf{f}'(\mathbf{x}^n)$  denotes the Fréchet derivative computed for the absorption parameter estimate  $\mathbf{x}^n$ . For the discretized domain,  $\Delta x_i = \mu_a(r_i) - \mu_a^n(r_i)$  and we have shown that a suitable approximation to the  $(l, i)$ -th element of the Fréchet derivative matrix is [4]

$$\frac{\partial \mathbf{f}_i(d_m)}{\partial x_i} = g(d_m, r_i) \phi_k(r_i) \left\{ -1 + \frac{\mu_a^n(r_i) + j\omega/c}{\mu_a^n(r_i) + \mu_s'(r_i)} \right\} A, \quad (10)$$

where  $A$  is the pixel area,  $l = M(k-1) + m$ , and  $g(d_m, r_i)$  is the Green's function for the frequency domain diffusion equation (1). After the  $n$ th scan,  $\mathbf{f}(\mathbf{x}^n)$  and  $\mathbf{f}'(\mathbf{x}^n)$  are calculated by computing  $g(d_m, r_i)$  and  $\phi_k(r_i)$  in (10) using a linear PDE solver for the diffusion equation (1) with the  $n$ th estimate of absorption coefficient,  $\mathbf{x}^n$ . These values are then used throughout the  $(n + 1)$ th scan, in which each pixel is updated in turn, and the new value  $\hat{x}_i$  is given by

$$\hat{x}_i = \arg \min_{\hat{x}_i \geq 0} \left[ \frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x}^n) - [\mathbf{f}'(\mathbf{x}^n)]_{*i}(\hat{x}_i - x_i^n)\|_{\Lambda}^2}{\alpha} + \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_i} b_{i-j} |\hat{x}_i - x_j|^p \right], \quad (11)$$

where  $[\mathbf{f}'(\mathbf{x}^n)]_{*i}$  is the  $i$ -th column of the Fréchet derivative matrix. The solution to (11) can be reduced to a linear half-interval search [5].

#### 4. Simulations

Simulations were used to assess the performance of the ICD/Born algorithm. Twelve sources and 12 detectors are located uniformly over the boundary of a  $8cm \times 8cm$  domain, as shown in Figure 1, and a modulation frequency of 200MHz is used. Random noise with a complex Gaussian distribution is added to the measured flux data (average snr = 30dB). The image is discretized on a  $33 \times 33$  grid, and  $\mu_s = 10.0cm^{-1}$ . The stopping criterion used was a fixed CPU time of 60 seconds (in order to allow comparison of the algorithms). Results are shown in Fig. 2. Comparison of the original image (top left) with that reconstructed using a modified DBIM [4] (top right) shows a high noise level and incorrect peak heights. For the GGMRF prior model, we used an eight-point neighborhood system, with  $b_{i-j} = (2\sqrt{2} + 4)^{-1}$  for nearest neighbors and  $b_{i-j} = (4\sqrt{2} + 4)^{-1}$  for diagonal neighbors. The MAP reconstruction computed with the ICD/Born algorithm for a Gaussian prior ( $p = 2, \sigma = 1.0 \times 10^{-3}$ ) is shown in the bottom left, and the reconstruction for  $p = 1.1$  and  $\sigma = 2.3 \times 10^{-4}$  is shown in the bottom right. The reconstruction using the Gaussian prior ( $p = 2$ ) reduces the background noise compared to that from the DBIM, but suffers from some edge smoothing. Sharper edges and good noise suppression are obtained for  $p = 1.1$ . Figure 3 shows the normalized root mean square error (NRMSE) in  $\mu_a$  versus CPU time. Note the eventual divergence of the DBIM due to insufficient regularization. Figure 3 also shows the cost function in (7) versus iteration for the ICD/Born method with  $p = 2.0$  and  $p = 1.1$ , showing that the method is stable and has reasonably fast convergence. A speedup by a factor of approximately 10-20 is obtained for the ICD/Born method over the DBIM [5].

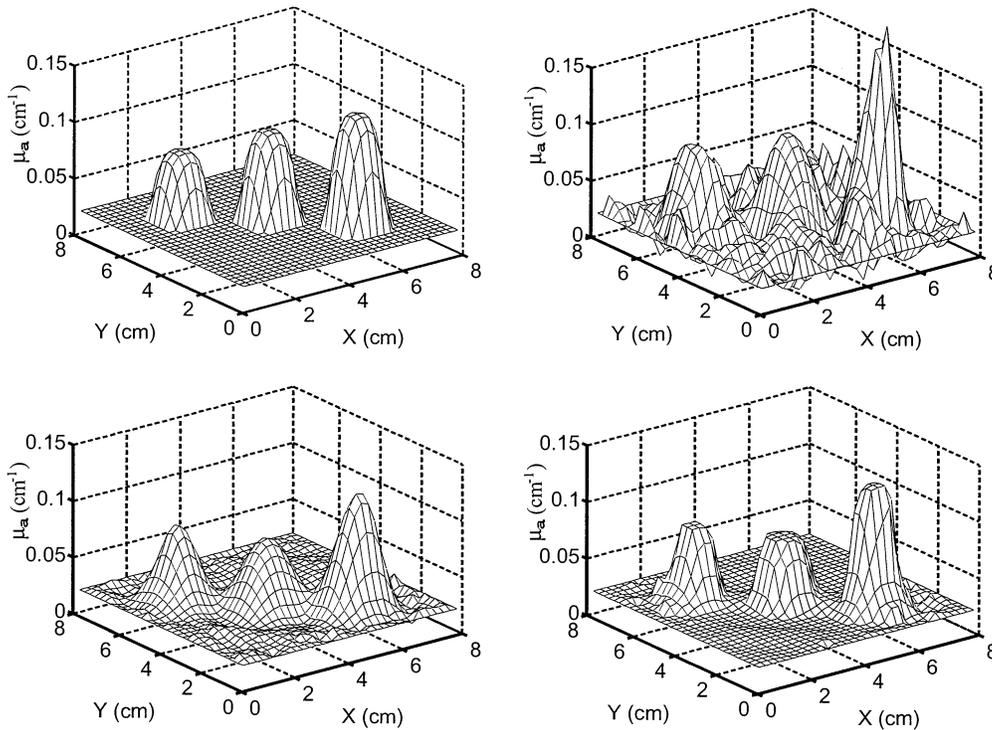


Figure 2: Original absorption image (top left), reconstruction by a modified DBIM (top right), reconstruction by the ICD/Born algorithm with a Gaussian prior ( $p = 2.0, \sigma = 1.00 \times 10^{-3}$ ) (bottom left), and reconstruction by the ICD/Born algorithm with a GGMRF prior ( $p = 1.1, \sigma = 2.31 \times 10^{-4}$ ) (bottom right).

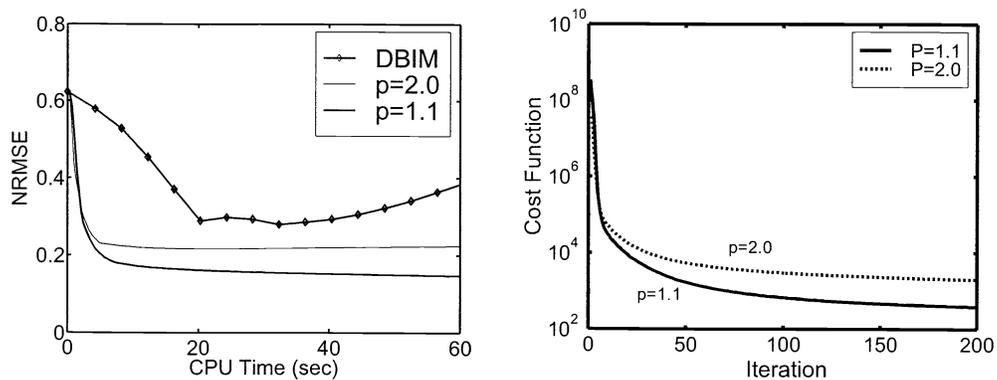


Figure 3: Normalized root mean square error (NRMSE) versus CPU time for the ICD/Born method (with  $p = 1.1$  and  $p = 2$ ) and the DBIM, and the cost function (log scale) versus iteration for the ICD/Born algorithm with  $p = 1.1$  and  $p = 2$ .

## 5. Multigrid Inversion

A realistic optical diffusion imaging problem is three-dimensional since the photon paths are not restricted to a plane. The formulation and results described above for two dimensions apply equally well in three dimensions. However, the computational size of the problem increases considerably when one moves from two dimensions to three. The three-dimensional problem has to be tackled however to address practical imaging problems. Despite this, most current work has considered only the two-dimensional case. Although the ICD/Born reconstruction method described above is efficient compared to previous DBIM-based approaches, it still presents a large computational burden for practical three-dimensional imaging. Multigrid techniques offer advantages in this regard since they tend to be computationally more efficient than fixed grid algorithms, and they tend to be more robust with respect to the value of hyperparameters [9].

We have developed a general multigrid optimization algorithm suitable for nonlinear image reconstruction problems, and applied it to the optical diffusion imaging problem [10,11]. This is in contrast to conventional multigrid algorithms which are formulated to solve differential or integro-differential equations [12-15]. In this algorithm, each iteration alternates a Born approximation step with a single cycle of a nonlinear multigrid algorithm. At each resolution of the multigrid algorithm the ICD method is used to optimize the solution.

The nonlinear multigrid algorithm is outlined as follows [11]. Let  $\mathbf{x}^{(0)} = \mathbf{x}$  denote the finest grid absorption image, and let  $\mathbf{x}^{(k)}$  be a coarser scale representation of  $\mathbf{x}^{(0)}$  with a grid spacing of  $2^k$ . Multigrid methods involve recursively moving between adjacent scales  $k$  and  $k+1$ , and can be described by defining, given an approximate solution  $\mathbf{x}^{(k)}$ , the computation of a solution at the next coarse grid level  $\mathbf{x}^{(k+1)}$  and use of this to improve the solution  $\mathbf{x}^{(k)}$ . In general,  $\mathbf{x}^{(k+1)}$  may be computed from  $\mathbf{x}^{(k)}$  by some linear transformation  $\mathbf{x}^{(k+1)} = \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}^{(k)}$  where  $\mathbb{I}_{(k)}^{(k+1)}$  is an  $(N/4^{k+1}) \times (N/4^k)$  decimation matrix. The corresponding linear interpolation matrices are denoted by  $\mathbb{I}_{(k+1)}^{(k)}$ .

Assume we need to minimize a cost functional  $c^{(k)}(\mathbf{x}^{(k)})$  at scale  $k$ , and that we have an initial solution  $\hat{\mathbf{x}}^{(k)}$  that approximately minimizes the cost functional, i.e.

$$\hat{\mathbf{x}}^{(k)} \approx \arg \min_{\mathbf{x}^{(k)}} \left\{ c^{(k)}(\mathbf{x}^{(k)}) \right\} . \quad (12)$$

The objective is to compute a solution at the next coarser grid,  $\hat{\mathbf{x}}^{(k+1)}$ , and then use this solution to improve or correct the fine grid solution. This fine-grid correction may be done using the formula

$$\hat{\mathbf{x}}^{(k)} \leftarrow \hat{\mathbf{x}}^{(k)} + \mathbb{I}_{(k+1)}^{(k)} \left( \hat{\mathbf{x}}^{(k+1)} - \mathbb{I}_{(k)}^{(k+1)} \hat{\mathbf{x}}^{(k)} \right) . \quad (13)$$

In order to compute the coarse grid solution,  $\hat{\mathbf{x}}^{(k+1)}$ , we must formulate a corresponding coarse grid optimization problem. To do this, we first choose a coarse grid cost functional,  $c^{(k+1)}(\mathbf{x}^{(k+1)})$ , which we believe to be a good approximation to  $c^{(k)}(\mathbf{x}^{(k)})$ . Of course the particular choice of this functional is very important and depends on the details of the problem being solved. However for now, simply assume that  $c^{(k+1)}(\mathbf{x}^{(k+1)})$  reasonably approximates the finer grid cost functional. To correct for possible discretization errors, we then solve an adjusted coarse scale optimization problem

$$\hat{\mathbf{x}}^{(k+1)} = \arg \min_{\mathbf{x}^{(k+1)} \geq \mathbf{0}} \left\{ c^{(k+1)}(\mathbf{x}^{(k+1)}) - \mathbf{r}^{(k+1)} \mathbf{x}^{(k+1)} \right\} , \quad (14)$$

where the row vector  $\mathbf{r}^{(k+1)}$  is the so-called residual term used to partially correct the approximation error between the coarse and fine grid cost functionals.

The question remains as to how we should choose the residual term  $\mathbf{r}^{(k+1)}$ . Ideally, we would like the following approximate equality to hold

$$c^{(k+1)}(\mathbf{x}^{(k+1)}) - \mathbf{r}^{(k+1)}\mathbf{x}^{(k+1)} \approx c^{(k)}\left(\hat{\mathbf{x}}^{(k)} + \mathbb{I}_{(k+1)}^{(k)}\left(\mathbf{x}^{(k+1)} - \mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}\right)\right). \quad (15)$$

In the general case, the difference between the lhs and rhs of (15) is not linear, so no choice of the row vector  $\mathbf{r}^{(k+1)}$  will achieve equality. However, we can choose  $\mathbf{r}^{(k+1)}$  to match the derivatives of the two sides when  $\mathbf{x}^{(k+1)} = \mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}$ . This condition results in the following important expression for  $\mathbf{r}^{(k+1)}$

$$\mathbf{r}^{(k+1)} = \nabla c^{(k+1)}(\mathbf{x})\Big|_{\mathbf{x}=\mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}} - \nabla c^{(k)}(\hat{\mathbf{x}}^{(k)})\mathbb{I}_{(k+1)}^{(k)}, \quad (16)$$

where  $\nabla c(\mathbf{x})$  denotes the row vector formed by the gradient of the functional  $c(\mathbf{x})$ .

There are three important observations to be made about (16). First, the expression holds for general choices of the cost functionals, and the interpolation and decimation operators. Second, if the cost functionals are strictly convex, then this choice of residual term ensures that an exact solution to the fine scale problem is also a fixed point to the coarse scale problem. To see this, observe that for an exact solution  $\nabla c^{(k)}(\hat{\mathbf{x}}^{(k)}) = 0$ , which implies that

$$\nabla\left(c^{(k+1)}(\mathbf{x}) - \mathbf{r}^{(k+1)}\mathbf{x}\right)\Big|_{\mathbf{x}=\mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}} = 0, \quad (17)$$

so that  $\hat{\mathbf{x}}^{(k+1)} = \mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}$  is a global minimum to (14). Hence the coarse grid correction (13) does not change an exact fine scale solution. This property is critical since it ensures that the exact fine grid solution is at least a fixed point to the final multigrid algorithm. Third, the interpolation matrix,  $\mathbb{I}_{(k+1)}^{(k)}$ , functions as a decimator in equation (16) since it is multiplied by the gradient vector from the left.

Multigrid optimization is implemented by recursively applying the two-grid optimization as described above. We have used two recursions known as V-cycle and full multigrid [9]. Each of these algorithms moves back and forth through coarse and fine resolution in a characteristic pattern as shown in Figure 4. Gradients are determined only at the finest grid. In the V-cycle algorithm, starting at the finest resolution, the two-grid algorithm described above is applied recursively, working down to the coarsest grid, and then working back to the finest grid, as shown in Figure 4(a). In the full multigrid scheme, the reconstruction starts from the coarsest grid. The resulting coarsest grid solution is interpolated to the next finer grid and used as the first approximation to the solution of the corresponding optimization problem, which is then solved by a multigrid V-cycle. This process is repeated, interpolating the solution from one level to the next as a first approximation and solving by a V-cycle, until the final solution is obtained at the finest level as shown in Figure 4(b).

For application to the optical diffusion imaging problem, we use for  $\mathbb{I}_{(k)}^{(k+1)}$  the separable extension of the one-dimensional decimation matrix

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}, \quad (18)$$

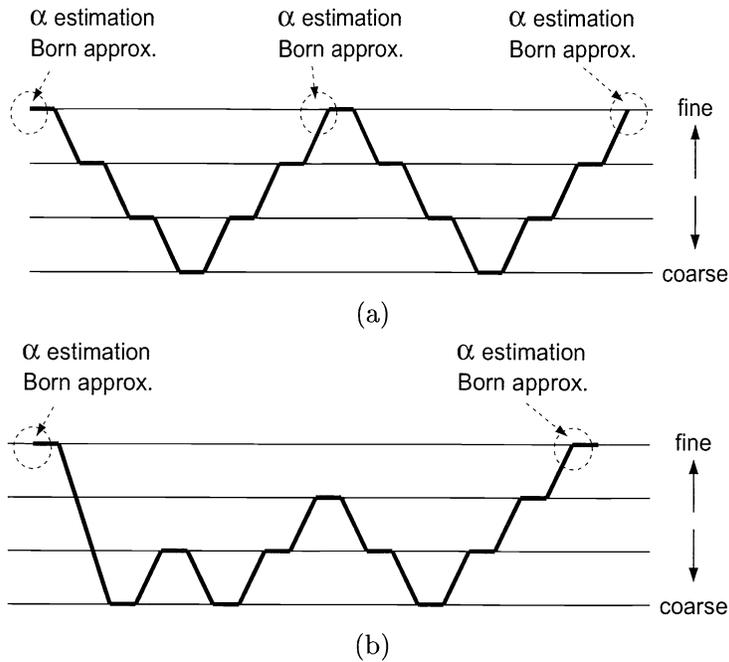


Figure 4: Multigrid inversion algorithms. Each iteration alternates a Born approximation step with a single cycle of a nonlinear multigrid algorithm. (a) V-cycle inversion algorithm, and (b) full multigrid inversion algorithm.

and the corresponding interpolation matrix

$$\mathbb{I}_{(k+1)}^{(k)} = 4 \left( \mathbb{I}_{(k)}^{(k+1)} \right)^T . \quad (19)$$

The coarse grid cost function and residual term are easily derived for the optical diffusion problem and are given in [11]. At each resolution, the ICD optimization method as described above is used to compute an approximate solution to the optimization. We now also consider  $\alpha$  to be unknown, so that the optimization problem (7) is replaced by

$$\hat{\mathbf{x}}_{MAP} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \min_{\alpha} \left\{ \frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\Lambda}^2}{\alpha} + P \log \alpha + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \right\} . \quad (20)$$

From (20), we can easily obtain the closed form expression for  $\alpha$

$$\alpha = \frac{1}{P} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\Lambda}^2 . \quad (21)$$

We achieve the MAP estimate by alternately maximizing with respect to  $\alpha$  and  $\mathbf{x}$  using the two equations

$$\hat{\alpha} = \frac{1}{P} \|\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}})\|_{\Lambda}^2 \quad (22)$$

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \left\{ \frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\Lambda}^2}{\hat{\alpha}} + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \right\} . \quad (23)$$

## 6. Multigrid Simulations

Some results of applying this algorithm to a  $129 \times 129$  image are described here. A larger image than that described in Section 4 is possible because of the increased speed of multigrid methods.

The true absorption image for the first example is shown in Fig. 5 (left). The absorption coefficient at each grid point, including the background, is considered unknown. The maximum absorption is  $0.15\text{cm}^{-1}$  and the background absorption is  $0.02\text{cm}^{-1}$ .  $\mu'_s$  is uniform with a value of  $10.0\text{cm}^{-1}$  and is considered known. Parameters of the simulation are the same as described in Section 4. The number of grid levels used in the multigrid algorithms is determined by decimating the finest grid image with a size of  $129 \times 129$  pixels until we obtain a grid of  $17 \times 17$  pixels (4 levels of resolution). A value of 1.1 was used for  $p$  in all the reconstructions. For each cycle of the ICD algorithm, we scan through the points in a new randomized order. A constant background with  $0.02\text{cm}^{-1}$  is used as the initial guess. Experimentation using the multigrid algorithm showed that the solution obtained and the convergence rate depended only weakly on the value of the GGMRF hyper-parameter  $\sigma$  used. The value  $\sigma = 0.04\text{cm}^{-1}$  actually best fits the statistics of the image and this value was used for the multigrid results shown here. For the fixed grid ICD/Born algorithm,  $\alpha$  was estimated by (22), but the fixed grid algorithm was used to solve (23).

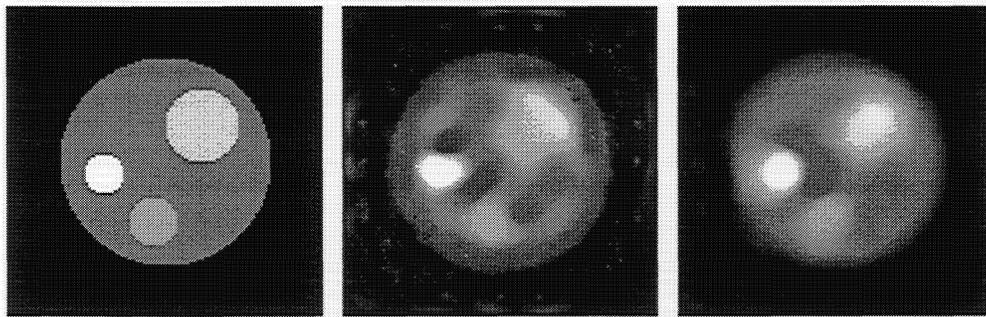


Figure 5: Gray scale of the absorption coefficient image (left), reconstruction by the fixed resolution ICD/Born algorithm (center), and reconstruction by a multigrid inversion algorithm (right). Black denotes  $\mu_a = 0.02\text{cm}^{-1}$  and white denotes  $\mu_a = 0.12\text{cm}^{-1}$ .

Convergence of the log posterior probability and the NRMSE for the full multigrid inversion algorithm and the fixed grid ICD/Born algorithm, as a function of CPU time, are shown in Figure 6. Note the significant computational saving of the multigrid algorithms over the fixed grid ICD/Born method. The full multigrid inversion algorithm converges at an early stage of the iterations. It was found that convergence of the fixed grid algorithm with  $\sigma = 0.04\text{cm}^{-1}$  was impractically slow, but using a value of  $\sigma = 0.1\text{cm}^{-1}$  gave better convergence (Fig. 6). The reconstructions using the fixed grid and full multigrid algorithms are shown in Figure 5. In this case, the full multigrid algorithm produces a more accurate reconstruction. The number of iterations and CPU times for the reconstructions shown in Figure 5 are listed in Table 1.

A variety of absorption cross-sections are shown in Figure 7, and their reconstructions using the full multigrid algorithm in Figure 8. In all cases  $\mu'_s(r) = 10.0\text{cm}^{-1}$ , the peak value of the absorption coefficient is  $0.08\text{cm}^{-1}$ , and the background is  $\mu_a(r) = 0.02\text{cm}^{-1}$ . The reconstructions are for 10 cycles of the full multigrid inversion algorithm using  $p = 1.1$  and  $\sigma = 0.02\text{cm}^{-1}$ . The NRMSE and CPU times

Table 1: Performance of the fixed grid and multigrid algorithms (Figure 5).

Method	$\sigma$	Iterations	CPU time (sec)
Fixed grid	0.1	1000	8900
Full multigrid	0.04	200	4100

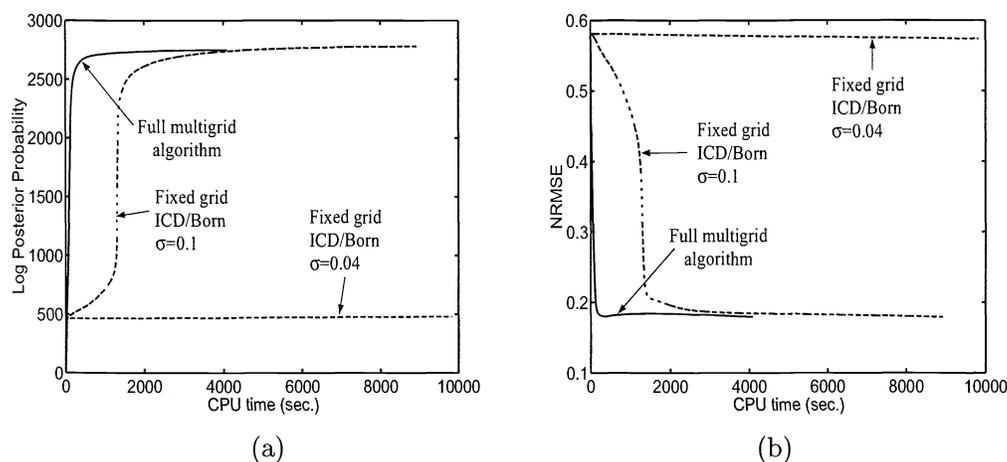


Figure 6: (a) Log posterior probability and (b) NRMSE as a function of CPU time.

are listed in Table 2. The reconstructions are quite accurate quantitatively and qualitatively, and have a small computational burden.

## 7. Conclusions

Optical imaging of highly scattering media with light offers a potentially powerful, flexible and inexpensive imaging modality. However, it represents a computationally demanding nonlinear image reconstruction problem. By formulating the problem in a Bayesian framework, and incorporating a multigrid algorithm with iterative coordinate descent optimization, we have been able substantially improve the quality of reconstructed images and increase the speed by at least a factor of 100 over conventional DBIM

Table 2: NRMSE and CPU time for the examples of Figure 8 after 10 cycles of the full multigrid inversion algorithm

	(a)	(b)	(c)	(d)	(e)	(f)
NRMSE	0.030	0.070	0.055	0.195	0.208	0.217
CPU time (sec)	221	224	236	239	234	223

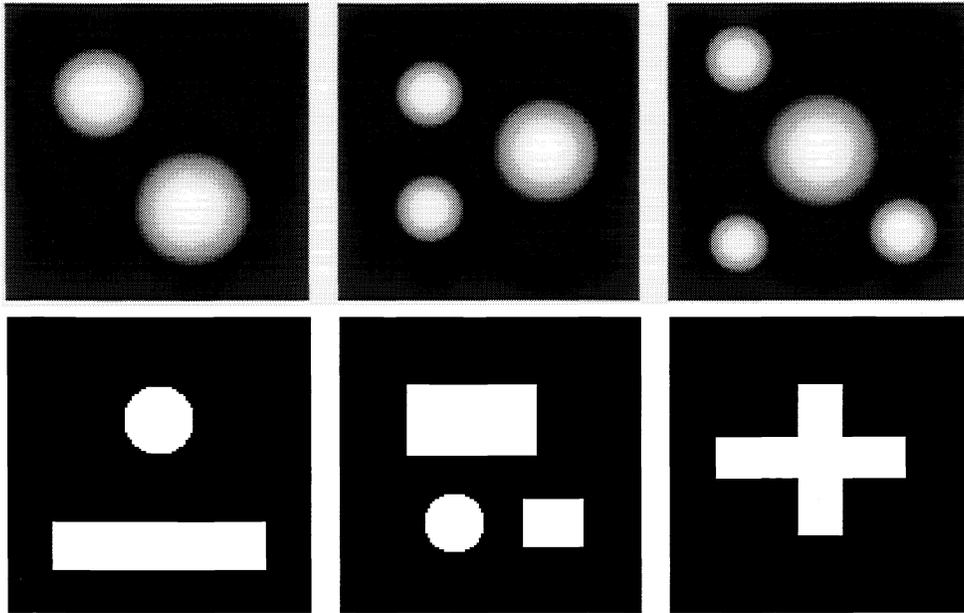


Figure 7: A variety of absorption image phantoms.

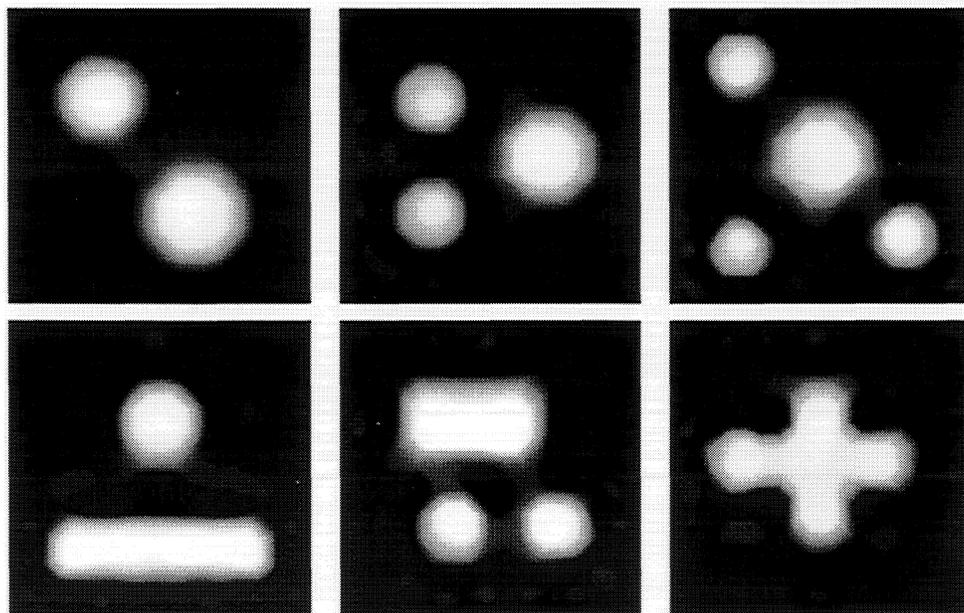


Figure 8: Reconstructions of the phantoms shown in Figure 7.

methods. These advances promise to open the door to practical three-dimensional optical imaging of highly scattering media.

Supported by the National Science Foundation (MIP-9707763 and CCR-0073357).

## References

1. M. S. Patterson, B. Chance, and B. Wilson: Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties. *Appl. Opt.* **28** (1989) 2331–2336.
2. Y. Yao, Y. Wang, Y. Pei, W. Zhu, and R. L. Barbour: Frequency domain optical imaging of absorption and scattering distributions by a Born iterative method. *J. Opt. Soc. Am. A* **14** (1997) 325–342.
3. J. C. Ye, R. P. Millane, K. J. Webb and T. J. Downar: Importance of the  $\nabla D$  term in frequency-resolved optical diffusion imaging. *Opt. Lett.* **23** (1998) 1423–1425.
4. 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. Opt. Soc. Am. A* **16** (1999) 1814–1826.
5. 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. Opt. Soc. Am. A* **16** (1999) 2400–2412.
6. C. A. Bouman and K. Sauer: A generalized Gaussian image model for edge-preserving map estimation. *IEEE Trans. Image Process.* **2** (1993) 296–310.
7. C. A. Bouman and K. Sauer: A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. Image Process.* **5** (1996) 480–492.
8. H. Carfantan, A. Mohammad-Djafari, and J. Idier: A single site update algorithm for nonlinear diffraction tomography. *Proc. of IEEE Int'l Conf. ASSP*, IEEE, New York, **4** (1997) 2837–2840.
9. J. C. Biggs: *A Multigrid Tutorial* SIAM, Philadelphia (1987).
10. J. C. Ye, C. A. Bouman, K. J. Webb and R. P. Millane: Nonlinear multigrid optimization for Bayesian diffusion tomography. *Proc. IEEE Int. Conf. Image Process.* (1999).
11. J. C. Ye, C. A. Bouman, K. J. Webb and R. P. Millane: Nonlinear multigrid algorithms for Bayesian optical diffusion tomography. *IEEE Trans. Image Process.* (2000) in press.
12. S. R. Fulton, P. E. Ciesielski, and W. H. Schubert: Multigrid methods for elliptic problems: A review. *J. Atmos. Sci.* **114** (1986) 943–959.
13. D. Terzopoulos: Image analysis using multigrid relaxation methods. *IEEE Trans. on Patt. Anal. and Machine Intell.* **PAMI-8** (1986) 129–139.
14. C. A. Bouman and K. Sauer: Nonlinear multigrid methods of optimization in Bayesian tomographic image reconstruction. *Proc. SPIE* **1766** (1992) 296–306.
15. S. F. McCormick and J. G. Wade: Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography. *Inverse Problems* **9** (1993) 697–713.