# Nonlinear Multigrid Algorithms for Bayesian Optical Diffusion Tomography

Jong Chul Ye, Charles A. Bouman, *Fellow, IEEE*, Kevin J. Webb, *Senior Member, IEEE*, and
Rick P. Millane, *Senior Member, IEEE*

***Abstract*—Optical diffusion tomography is a technique for imaging a highly scattering medium using measurements of transmitted modulated light. Reconstruction of the spatial distribution of the optical properties of the medium from such data is a difficult nonlinear inverse problem. Bayesian approaches are effective, but are computationally expensive, especially for three-dimensional (3-D) imaging. This paper presents a general nonlinear multigrid optimization technique suitable for reducing the computational burden in a range of nonquadratic optimization problems. This multigrid method is applied to compute the maximum *a posteriori* (MAP) estimate of the reconstructed image in the optical diffusion tomography problem. The proposed multigrid approach both dramatically reduces the required computation and improves the reconstructed image quality.**

***Index Terms*—Bayesian image reconstruction, multiresolution image reconstruction, nonlinear multigrid optimization, optical diffusion tomography.**

## I. INTRODUCTION

OPTICAL diffusion imaging is a technique for reconstructing the optical parameters in highly scattering media such as tissue, polymer composites, sea ice, and aerosols, based on measurements of the scattered and attenuated optical energy. For tissue imaging, this technique presents significantly lower health risks as compared to X-ray imaging, and is instrumentally much less expensive than X-ray CT or MRI. Moreover, the potential of optical diffusion imaging has been successfully demonstrated in biomedical applications [1]. However, a major difficulty with this approach is that the relationship between the unknown scattering and attenuation coefficients and the optical measurements is highly nonlinear and described by a partial differential equation; so reconstruction poses a challenging nonlinear inverse problem.

Inversion approaches for optical diffusion tomography based on the Born or Rytov approximations [2] produce significant errors in the reconstruction for realistic material parameters due to linearization errors of the forward model. To overcome these drawbacks, iterative techniques have been investigated. Usually, the Newton–Raphson (NR) method has been used with a Levenberg–Marquardt procedure. A Levenberg–Marquardt method for a variational formulation of the diffusion equation has been applied to the time-domain problem as well as the frequency-domain problem [3], [4]. However, the constraint used in these approaches [3], [4], which imposes a penalty on the $L_2$ norm of the new update at each iteration, tends to over-smooth edges in the image or produce excessively noisy images, depending on a control parameter value. A fundamental drawback of these methods is that the $L_2$ penalty term for the new update is not a form of regularization in the Tikhonov sense [5], but is instead a "trust region" constraint designed to insure monotone convergence of the optimization criterion [6], [7].

The artifacts due to poor regularization can be reduced by incorporation of prior information using a Bayesian framework. Recently, Bayesian (and other regularization) methods have been applied to nonlinear inverse problems such as microwave imaging and optical diffusion imaging [3], [8]–[12]. The individual approaches have differed both in terms of the prior model (or stabilizing functional) used and the optimization algorithms employed to compute the reconstruction. For example, Paulsen and Jiang added a quadratic regularization term to their previous formulation [3] to stabilize the reconstruction [8]. In this case, each iteration of the optimization performed a linearization (similar to the Born approximation) followed by a full matrix inversion to solve the linearized problem. The computational complexity of this method is very high since $O(N^3)$ complex multiplications are required at each iteration, where $N$ is the number of image pixels. Saquib, Hanson, and Cunningham proposed a more computationally efficient algorithm for the time-domain diffusion problem in which each iteration alternates a linearization step with a single step of a conjugate gradient algorithm [9], [13]. This work drew on the concept of adjoint differentiation [14] for efficient computation of the gradient. Arridge and Schweiger applied this gradient method to frequency domain optical diffusion imaging [10], [15]. However, the complexity of the line search [7] required in a conjugate gradient algorithm is an important factor in the total computational burden for both approaches.

More recently, Bayesian approaches based on iterative coordinate descent (ICD) optimization have been investigated [11], [12]. The ICD method is a fast implementation of the Gauss–Seidel method that is well-suited for tomography applications [16], [17]. In particular, we have developed an ICD/Born method that provides high quality reconstructions and is computationally efficient when compared to the conventional iterative Born approximation methods [12]. Since

---

Manuscript received February 22, 1999; revised November 14, 2000. This work was supported by a Purdue Research Foundation Fellowship and the NSF under Grants MIP-9707763 and CCR-0073357. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. William Clem Karl.

J. C. Ye, C. A. Bouman, and K. J. Webb are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA (e-mail: bouman@purdue.edu).

R. P. Millane is with the Computational Science and Engineering Program, Purdue University, West Lafayette IN 47907-1160 USA.

Publisher Item Identifier S 1057-7149(01)01652-9.

1057–7149/01$10.00 © 2001 IEEE

the photon paths are not restricted to a plane, practical and accurate optical diffusion imaging will require inversion for three-dimensional (3-D) images, with a concomitant increase in the number of unknowns. The computational complexity of the ICD/Born method is still prohibitive for such problems.

Multigrid algorithms are a specific form of multiresolution algorithm that can be used to reduce the computational requirements of large numerical problems [18]–[20]. These algorithms work by recursively moving between different resolutions, thereby propagating information between coarse and fine scales. Multigrid methods have been primarily used for solving partial differential equations [21], but more recently they have been applied to a variety of imaging problems such as image analysis [22], [23] and anisotropic diffusion [24].

Perhaps surprisingly, multigrid algorithms have not been widely applied in tomography problems. In earlier work, Bouman and Sauer [25] used multigrid algorithms to solve the nonquadratic optimization problems resulting from projection tomography applications such as computed tomography (CT), and photon emission tomography (PET). While this formulation used nonlinear multigrid, it was based on a conventional nonlinear multigrid PDE solver. Other research by McCormick and Wade [26] used multigrid algorithms for impedance tomography problems. This work linearized the impedance tomography problem, and therefore used a standard linear multigrid equation solver. Bhatia *et al.* [27] and Zhu *et al.* [28] used wavelet methods to solve linear or linearized tomography problems.

In this paper, we develop a multigrid optimization method suitable for solving general nonquadratic optimization problems; and we apply this method to the problem of optical diffusion tomography. Multigrid algorithms are well suited to this problem for three reasons. First, our simulations indicate that our multigrid algorithm converges much faster than fixed grid algorithms. This is particularly important for the optical diffusion tomography problem since it is inherently 3-D. Second, multigrid algorithms are well suited for implementation of positivity constraints because the optimization at each grid resolution is done in the space-domain where positivity constraints are easily enforced. In general, positivity can be important for improving reconstruction quality, particularly when the problem is underdetermined. In the optical diffusion tomography problem the physical parameters being inverted can have only positive values. Finally, multigrid algorithms tend to better avoid local minima, or tend to find a better local minimum, in the functional being optimized. Since the diffusion tomography problem results in a nonconvex optimization problem, this robustness to local minima helps insure that a good solution is reached.

A key innovation of our work is the direct formulation of the multigrid algorithm in an optimization framework. Historically, multigrid techniques were developed for solving linear or nonlinear elliptic PDEs [18]. While they can be used to solve optimization problems, this is generally done by differentiating the cost function, and using multigrid algorithms to solve the resulting equation. In contrast, we have derived expressions for the direct application of multigrid methods to optimization problems. This approach greatly simplifies the application of multigrid to our problem.

Section II describes the forward diffusion equation model and reviews the Bayesian cost function in this context. Alternating estimation of the data term noise variance and the updated image is then described as a means to vary the degree of regularization and improve convergence. The multigrid algorithm we develop for nonquadratic optimization is presented in Section III. Section IV presents a complexity analysis of the multigrid and fixed grid algorithms. In Section V we present the results of simulations using our multigrid method in comparison with a fixed grid algorithm. Concluding remarks are made in Section VI.

## II. BAYESIAN FRAMEWORK FOR OPTICAL DIFFUSION TOMOGRAPHY

In this section, we develop the Bayesian framework that we use for reconstructing the material parameters of highly scattering media from measurements of scattered light. In Section II-A we develop the forward model based on the diffusion equation, and in Section II-B, we use this model to formulate the maximum *a posteriori* (MAP) optimization problem.

### A. Forward Model for Optical Diffusion Tomography

In a highly scattering medium, it is useful to look only at the intensity of the electromagnetic wave. Here, photons are treated as particles which elastically scatter through the random medium. The theoretical framework for this model is Boltzmann transport theory [29], which applies conservation of energy for the photon density scatter and source mechanisms. A common approximation to the Boltzmann transport equation is the diffusion equation [29], [30], which assumes that the flux has a weak angular dependence, that all photons travel at the same speed, that the sources are isotropic, and that the photon current density changes slowly with time, relative to the mean collision time [29]. The diffusion approximation is accurate in soft tissue over the 650–1300 nm wavelength range where scatter dominates absorption [31]–[33], and provides a computationally tractable forward model for tissue imaging.

Let the scalar quantity $N_k(r)$ be the photon density (with the dimensions of energy per unit volume) at position $r \in \Omega$ due to a point source of light at position $s_k \in \Omega$, where $\Omega$ is the domain of interest. Then, the photon flux is defined as $\psi_k(r, t) = cN_k(r)$, with $c$ being the speed of light in the medium. The photon flux $\psi_k(r, t)$, which describes the optical power density as a function of position and time, satisfies the time domain diffusion equation

$$\begin{aligned}\frac{1}{c}\frac{\partial}{\partial t}\psi_k(r, t) - \nabla \cdot (D(r)\nabla\psi_k(r, t)) + \mu_a(r)\psi_k(r, t)\\ = S(t)\delta(r - s_k) \qquad (1)\end{aligned}$$

where $S(t)$ is the time varying photon source density, and $D(r)$ is the diffusion constant given by

$$D(r) = \frac{1}{3(\mu_a(r) + \mu_s'(r))} \qquad (2)$$

with $\mu_a(r)$ the absorption coefficient, and $\mu_s'(r)$ the reduced scattering coefficient. The reduced scattering coefficient is de-

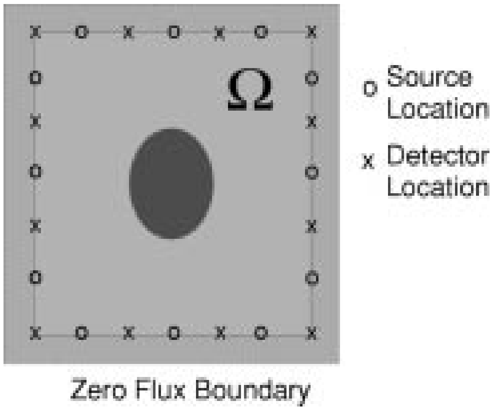


Fig. 1. Geometry used for simulation of optical diffusion measurements. The 12 detectors and 12 sources are uniformly spaced around the perimeter of the object.

fined by $\mu_s'(r) = (1 - g)\mu_s(r)$ where $\mu_s(r)$ is the scattering coefficient and $g$ is the mean cosine of the scattering angle.

Practical systems based on time domain measurements have been implemented [34], [31], but these systems tend to be expensive and noise sensitive. In order to circumvent these problems, we adopt a frequency domain approach to the optical diffusion problem [35], [3]. To do this, we assume that the light source is amplitude modulated at a fixed angular frequency $\omega (\neq 0)$, so that $S(t) = \mathrm{Re}[1 + \beta \exp(-j\omega t)]$, where $\beta$ is the modulation depth. At the detector, the complex modulation envelope is then measured by demodulating the in-phase and quadrature components of the measured sinusoidal signal, which is proportional to $\psi_k(r, t)$. This technique allows low noise narrow-band heterodyne detection of the modulation envelope $\psi_k(r, t)$ [36], which is denoted as the complex quantity $\phi_k(r)$. By taking the Fourier transform of (1), the partial differential equation that governs the complex modulation envelope, $\phi_k(r)$, becomes

$$\nabla \cdot (D(r)\nabla \phi_k(r)) + (-\mu_a(r) + j\omega/c)\phi_k(r) = -\beta\delta(r - s_k). \quad (3)$$

In the frequency domain imaging approach, (3) is used as the forward model, and the energy measured by a detector is then proportional to the photon current $\mathbf{J} = -D\nabla\phi(r)$ [29].

Fig. 1 shows a two-dimensional (2-D) imaging domain with interspersed source and detector points uniformly distributed around the boundary. We will use this as a representative experimental scenario for optical diffusion tomography. The region to be imaged is denoted by $\Omega$ and is surrounded by $K$ point sources at positions $s_k \in \Omega$ and $M$ detectors at positions $d_m \in \Omega$. In general one could image both the absorption and scattering coefficients as a function of position. However, here we consider the absorption imaging problem where we determine the values of $\mu_a(r)$ from the measured values of $\phi_k(d_m)$, while assuming $\mu_s'(r)$ known.

Using the same notation as in [12], measurements of the complex envelope $\phi_k(d_m)$ for source $k$ and detector $m$ are denoted by $y_{km}$. We also organize these measurements as a single column vector $\mathbf{y}$ of length $P = KM$ where $K$ denotes the number of sources, and $M$ denotes the number of detectors, i.e.,

$$\mathbf{y} = [y_{11}, y_{12}, \cdots, y_{1M}, y_{21}, \cdots, y_{KM}]^T. \quad (4)$$

The domain $\Omega$ is discretized into $N$ pixels, where the position of the $i$th pixel is denoted by $r_i$ for $1 \leq i \leq N$. The set of unknown absorption coefficients is denoted by the vector $\mathbf{x}$, where

$$\mathbf{x} = [\mu_a(r_1), \cdots, \mu_a(r_N)]^T. \quad (5)$$

In order to formulate this problem in a Bayesian framework, we require the data likelihood $p(\mathbf{y}|\mathbf{x})$. With the detectors operating at a sufficiently low temperature, photon detection can be modeled using shot noise statistics [12], which has its origin in Poisson statistics [36]. With $\phi_k(d_m)$ sufficiently large, i.e., with an adequate number of detected photons, the measurements are independent complex Gaussian random variables, and the data likelihood is given by [12]

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(\pi\alpha)^P |\Lambda|^{-1}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_\Lambda^2}{\alpha}\right] \quad (6)$$

where $\alpha$ is the parameter related to the noise variance, $\Lambda$ is the diagonal covariance matrix, $\|\mathbf{w}\|_\Lambda^2 = \mathbf{w}^H \Lambda \mathbf{w}$, 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 $\mathbf{x}$. More specifically, $\mathbf{f}(\mathbf{x}) = E[\mathbf{y}]$, where $E[\,\cdot\,]$ denotes expectation, is given by

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

For our problem, the measurements are statistically independent with the variance of each measurement equal to its mean; so $\Lambda$ is diagonal. For our simulations, we make the assumption that

$$\Lambda_{ii} \simeq \frac{1}{|y_{km}|}, \qquad \text{where } i = M(k-1) + m. \quad (7)$$

This approximation results from the assumption that the DC and modulated light undergo the same loss, which is approximately true for low modulation frequencies [12].

### *B. Formulation of the Bayesian Optimization Problem*

Bayesian methods provide a natural framework for incorporating prior information about the behavior of the unknown quantity $\mathbf{x}$. The MAP estimate of $\mathbf{x}$ given the measurement vector $\mathbf{y}$ is

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\max_{\mathbf{x} \geq 0} \{\log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x})\} \quad (8)$$

where $p(\mathbf{x})$ is the prior density for the image and maximization over $\mathbf{x} \geq 0$ enforces the required positivity constraint, i.e., that $\mu_a > 0$, as required for the physical problem. As in [12], we use the generalized Gaussian Markov random field (GGMRF) prior model [37]

$$p(\mathbf{x}) = \frac{1}{\sigma^N z(p)} \exp\left[-\frac{1}{p\sigma^p} \sum_{\{i, j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p\right] \quad (9)$$

where $\sigma$ is a normalization hyperparameter and $1 \leq p \leq 2$ controls the degree of edge smoothness, with $p = 2$ corresponding to the Gaussian case. This prior model enforces smoothness in the solution while preserving sharp edge transitions.

We adaptively estimate $\alpha$ during the reconstruction procedure. Initially, the estimated value of $\alpha$ is 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 $\alpha$ decreases, making the data term more important and consequently reducing the relative importance of the regularization term. We have found that this progression of $\alpha$ from large to small values increases the robustness of convergence to the minimum. 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.

If we consider $\alpha$ unknown, referring to (6) and (9), the optimization problem (8) can be re-written as

$$\arg\max_{\mathbf{x}\geq\mathbf{0}}\max_{\alpha}\left\{-\frac{1}{\alpha}\,||\mathbf{y}-\mathbf{f}(\mathbf{x})||_\Lambda^2 - P\log\alpha - \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p + \text{const}\right\}. \tag{10}$$

Viewing (10) as a cost function, and setting the derivative with respect to $\alpha$ equal to zero, we obtain the closed form expression

$$\alpha = \frac{1}{P}\,||\mathbf{y}-\mathbf{f}(\mathbf{x})||_\Lambda^2. \tag{11}$$

By substituting (11) into (10), the optimization problem (10) is converted into

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\geq\mathbf{0}}\left\{-P - P\log\left(\frac{1}{P}\,||\mathbf{y}-\mathbf{f}(\mathbf{x})||_\Lambda^2\right) - \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p\right\} \tag{12}$$

where $\hat{\mathbf{x}}$ is an estimate of the unknown $\mathbf{x}$. After neglecting constant terms, we can define the log posterior probability $l(\mathbf{x})$ as

$$l(\mathbf{x}) = -P\log||\mathbf{y}-\mathbf{f}(\mathbf{x})||_\Lambda^2 - \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p. \tag{13}$$

The log posterior probability (13) is used as a criterion for the convergence in our experimental results. For computational simplicity, we maximize $l(\mathbf{x})$ by alternately maximizing with respect to $\alpha$ and $\mathbf{x}$ using the following two equations:

$$\hat{\alpha} = \frac{1}{P}\,||\mathbf{y}-\mathbf{f}(\hat{\mathbf{x}})||_\Lambda^2 \tag{14}$$

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\geq\mathbf{0}}\left\{-\frac{1}{\hat{\alpha}}\,||\mathbf{y}-\mathbf{f}(\mathbf{x})||_\Lambda^2 - \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p\right\}. \tag{15}$$

Equation (14) is a straightforward computation, but (15) is a computationally expensive optimization problem, especially for large images $\mathbf{x}$. To circumvent this problem, we employ multigrid optimization algorithms to efficiently compute (15). At the beginning of each multigrid iteration the nonlinear functional $\mathbf{f}(\mathbf{x})$ is first linearized using a Taylor series expansion as

```
main {
1. Initialize $\hat{\mathbf{x}}$ with a background absorption coefficient estimate.
2. Repeat until converged: {
  (a) $\hat{\alpha} \leftarrow \frac{1}{P}||\mathbf{y}-\mathbf{f}(\hat{\mathbf{x}})||_\Lambda$ .
  (b) Compute the Fréchet derivative $\mathbf{f}'(\hat{\mathbf{x}})$ using (34), (35), (36).
  (c) Compute the following:

          $\mathbf{z} \leftarrow \mathbf{y}-\mathbf{f}(\hat{\mathbf{x}})+\mathbf{f}'(\hat{\mathbf{x}})\hat{\mathbf{x}}$
          $\mathbf{A} \leftarrow \mathbf{f}'(\hat{\mathbf{x}})$

  (d) Apply a multigrid optimization algorithm to minimize (17).

          $\hat{\mathbf{x}} \leftarrow \text{MultigridV}(\hat{\mathbf{x}}, \mathbf{r}=\mathbf{0}, \mathbf{z}, \mathbf{A}, k=0)$
          or
          $\hat{\mathbf{x}} \leftarrow \text{FMG}(\hat{\mathbf{x}}, \mathbf{r}=\mathbf{0}, \mathbf{z}, \mathbf{A}, k=0)$

  }
3. Stop.
}
```

Fig. 2. Pseudo-code specification for the optimization procedure. Each iteration of the procedure estimates the parameter $\hat{\alpha}$, recomputes the Born approximation, and then applies the multigrid optimization of Section III.

$$||\mathbf{y}-\mathbf{f}(\mathbf{x})||_\Lambda^2 \simeq ||\mathbf{y}-\mathbf{f}(\hat{\mathbf{x}})-\mathbf{f}'(\hat{\mathbf{x}})\Delta\mathbf{x}||_\Lambda^2 \tag{16}$$

where $\Delta\mathbf{x} = \mathbf{x}-\hat{\mathbf{x}}$, and $\mathbf{f}'(\hat{\mathbf{x}})$ represents the Fréchet derivative of $\mathbf{f}(\,\cdot\,)$ at $\hat{\mathbf{x}}$. Note that the Taylor series expansion of (16) turns out to be exactly the same as would result from a Born approximation [38]–[40]. The details of how the matrix $\mathbf{f}'(\hat{\mathbf{x}})$ is computed for this problem are given in Appendix A. Using (16), an approximate cost function for the original problem (15) is

$$c(\mathbf{x}) = \frac{1}{\hat{\alpha}}\,||\mathbf{z}-\mathbf{A}\mathbf{x}||_\Lambda^2 + \frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p \tag{17}$$

where

$$\begin{aligned}\mathbf{A} &= \mathbf{f}'(\hat{\mathbf{x}})\\ \mathbf{z} &= \mathbf{y}-\mathbf{f}(\hat{\mathbf{x}})+\mathbf{f}'(\hat{\mathbf{x}})\hat{\mathbf{x}}.\end{aligned}$$

Our overall strategy for the optimization of (10) is listed in the pseudo-code of Fig. 2 and is illustrated in Fig. 3. Each iteration of our algorithm starts with an update of $\alpha$ using (14), followed by a new linearization (17). This results in a nonlinear 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.

## III. NONLINEAR MULTIGRID INVERSION ALGORITHM

In this section, we derive a general algorithm for multigrid optimization of any functional, and we also derive the specific expressions for optimization of the cost functional in (17). Our approach is unique because it is formulated directly in an optimization framework. This is in contrast to conventional multigrid algorithms which are formulated to solve differential or integro-differential equations [19], [21], [22], [25], [26]. To derive

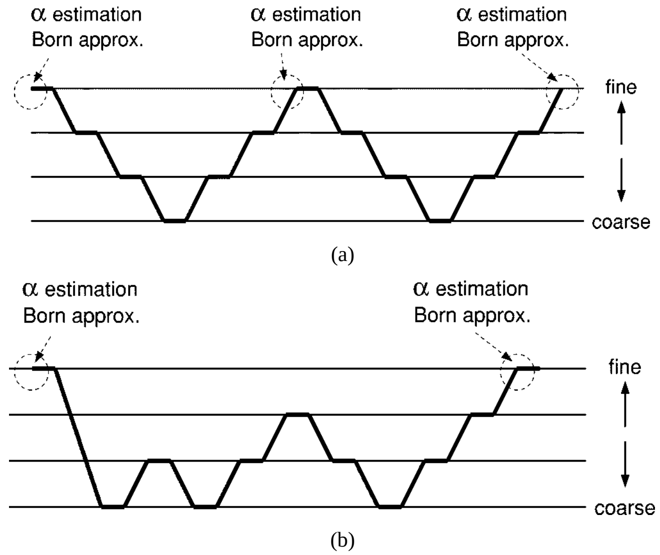


Fig. 3. Multigrid inversion algorithms. Each iteration alternates a Born approximation step with a single iteration of a nonlinear multigrid algorithm. (a) V-cycle inversion algorithm, and (b) full multigrid inversion algorithm.

our method, we start with the two-grid case, and then generalize this solution using standard recursions for the V-cycle and full multigrid cases [41].

### *A. Two Grid Algorithm*

For the two grid algorithm, we first consider optimization without the positivity constraint. We then discuss the addition of the positivity constraint in Section III-B.

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 sampling period of $2^k$ times the finest grid sampling period. 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 for the 2-D case. 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 $k$. Also assume that we have an initial solution $\hat{\mathbf{x}}^{(k)}$ which approximately minimizes the cost functional, i.e.,

$$\hat{\mathbf{x}}^{(k)} \approx \arg\min_{\mathbf{x}^{(k)}} \left\{ c^{(k)}\left(\mathbf{x}^{(k)}\right) \right\}. \tag{18}$$

Our objective is to compute the 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

$$\hat{\mathbf{x}}^{(k)} \leftarrow \hat{\mathbf{x}}^{(k)} + \mathbb{I}_{(k+1)}^{(k)}\left(\hat{\mathbf{x}}^{(k+1)} - \mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}\right). \tag{19}$$

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

$$\hat{\mathbf{x}}^{(k+1)} = \arg\min_{\mathbf{x}^{(k+1)}} \left\{ c^{(k+1)}\left(\mathbf{x}^{(k+1)}\right) - \mathbf{r}^{(k+1)}\mathbf{x}^{(k+1)} \right\} \tag{20}$$

where $\mathbf{r}^{(k+1)}$ is a constant row vector which may be used to adjust for errors in the cost function. This row vector is equivalent to the so-called residual term used to partially correct errors between coarse and fine grids in conventional multigrid [41].

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 for all values of $\mathbf{x}^{(k+1)}$:

$$\begin{aligned} &c^{(k+1)}\left(\mathbf{x}^{(k+1)}\right) - \mathbf{r}^{(k+1)}\mathbf{x}^{(k+1)} + \mathrm{const} \\ &\quad \approx c^{(k)}\left(\hat{\mathbf{x}}^{(k)} + \mathbb{I}_{(k+1)}^{(k)}\left(\mathbf{x}^{(k+1)} - \mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}\right)\right). \end{aligned} \tag{21}$$

The left hand side of (21) is the corrected coarse scale cost function, while the right hand side is the fine grid cost function evaluated using the corrected result of (19). If these two functions are equal, within a constant, then their minimum will occur for the same value of $\mathbf{x}^{(k+1)}$.

In the general case, the difference between the left and right hand sides of (21) is not linear, so no choice of the row vector $\mathbf{r}^{(k+1)}$ can 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 key expression for $\mathbf{r}^{(k+1)}$

$$\mathbf{r}^{(k+1)} = \nabla c^{(k+1)}(\mathbf{x})\Big|_{\mathbf{x}=\mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}} - \nabla c^{(k)}\left(\hat{\mathbf{x}}^{(k)}\right)\mathbb{I}_{(k+1)}^{(k)} \tag{22}$$

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

There are a number of observations to be made about (22). The expression holds for general choices of the cost functionals, the interpolating operators, and the decimating operators. It is interesting to note that the interpolation matrix, $\mathbb{I}_{(k+1)}^{(k)}$, actually functions as a decimation operator in equation (22) because it is being multiplied by the gradient vector from the left. Perhaps the most important observation is that the exact solution to (18) is a fixed point to this two grid update procedure. More precisely, the following theorem is proved in Appendix B.

*Theorem:* Let $c^{(k)}(\mathbf{x}^{(k)})$ and $c^{(k+1)}(\mathbf{x}^{(k+1)})$ be strictly convex continuously differentiable functionals on $\mathbb{R}^N$ and $\mathbb{R}^{N'}$, respectively, where $N' < N$, and let $\mathbf{x}^{*(k)} \in \mathbb{R}^N$ be the global minimum of $c^{(k)}(\,\cdot\,)$. Furthermore, consider the two grid update formed by applying (20) followed by (19) using the residual calculation of (22). Then the exact solution, $\mathbf{x}^{*(k)}$, is a fixed point of the two grid update.

### *B. Recursive Formulations of Multigrid Inversion*

Multigrid optimization is implemented by recursively applying the two-grid update of the previous section. In particular, we use the two recursions known as V-cycle and full multigrid [41]. The pseudo-code recursions for V-cycle (*MultigridV*) and full multigrid (*FMG*) with the use of positivity constraints are shown in Figs. 4 and 5. Each of these algorithms moves back

```
MultigridV($\mathbf{x}, \mathbf{r}, k$) {
1. Apply $\nu_1$ fixed grid iterations to compute:
       $\mathbf{x} \leftarrow\approx \arg\min_{\mathbf{x}\geq 0} \left\{ c^{(k)}(\mathbf{x}) - \mathbf{r}\mathbf{x} \right\}$
2. If $k$ is the coarsest desired grid, Return($\mathbf{x}$).
3. Compute the following:
       $\mathbf{x}_{dec} \leftarrow \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}$
       $\mathbf{r}^{(k+1)} \leftarrow \nabla c^{(k+1)}(\mathbf{x}_{dec}) - \nabla c^{(k)}(\mathbf{x}) \mathbb{I}_{(k+1)}^{(k)}$
4. $\mathbf{x}^{(k+1)} \leftarrow$ MultigridV($\mathbf{x}_{dec}, \mathbf{r}^{(k+1)}, k+1$)
5. Perform coarse grid correction:
       $\mathbf{x} \leftarrow \mathbf{x} + \mathbb{I}_{(k+1)}^{(k)} (\mathbf{x}^{(k+1)} - \mathbf{x}_{dec})$
       $\mathbf{x} \leftarrow \max\{\mathbf{x}, 0\}$
6. Apply $\nu_2$ fixed grid iterations to compute:
       $\mathbf{x} \leftarrow\approx \arg\min_{\mathbf{x}\geq 0} \left\{ c^{(k)}(\mathbf{x}) - \mathbf{r}\mathbf{x} \right\}$
7. Return($\mathbf{x}$)
}
```

Fig. 4. Pseudo-code specification of the proposed multigrid optimization method using the V-cycle recursion and a positivity constraint on the solution $\mathbf{x}$.

```
FMG($\mathbf{x}, \mathbf{r}, k$) {
1. If $k$ is the coarsest grid, go to 5.
2. Compute the following:
       $\mathbf{x}_{dec} \leftarrow \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}$
       $\mathbf{r}^{(k+1)} \leftarrow \nabla c^{(k+1)}(\mathbf{x}_{dec}) - \nabla c^{(k)}(\mathbf{x}) \mathbb{I}_{(k+1)}^{(k)}$
3. $\mathbf{x}^{(k+1)} \leftarrow$FMG($\mathbf{x}_{dec}, \mathbf{r}^{(k+1)}, k+1$)
4. Perform coarse grid correction
       $\mathbf{x} \leftarrow \mathbf{x} + \mathbb{I}_{(k+1)}^{(k)} (\mathbf{x}^{(k+1)} - \mathbf{x}_{dec})$
       $\mathbf{x} \leftarrow \max\{\mathbf{x}, 0\}$
5. $\mathbf{x} \leftarrow$MultigridV($\mathbf{x}, \mathbf{r}, k$).
6. Return($\mathbf{x}$)
}
```

Fig. 5. Pseudo-code specification for the full multigrid optimization algorithm using a positivity constraint on the solution $\mathbf{x}$.

```
MultigridV($\mathbf{x}, \mathbf{r}, \mathbf{z}, \mathbf{A}, k$)
1. Apply $\nu_1$ ICD iterations with initial condition $\mathbf{x}$ and compute:
       $\mathbf{x} \leftarrow\approx \arg\min_{\mathbf{x}\geq 0} \left\{ ||\mathbf{z} - \mathbf{A}\mathbf{x}||_{\Lambda}^2 + \frac{4^k \hat{\alpha}}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| \frac{x_i^{(k)} - x_j^{(k)}}{2^k} \right|^p - \mathbf{r}\mathbf{x} \right\}$
2. If $k$ is the coarsest desired grid, Return($\mathbf{x}$).
3. Compute the following:
       $\mathbf{x}_{dec} \leftarrow \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}$
       $\mathbf{A}^{(k+1)} \leftarrow \mathbf{A}\mathbb{I}_{(k+1)}^{(k)}$ (Equation (26))
       $\mathbf{z}^{(k+1)} \leftarrow \mathbf{z} - \mathbf{A}\left(I - \mathbb{I}_{(k+1)}^{(k)} \mathbb{I}_{(k)}^{(k+1)}\right) \mathbf{x}^{(k)}$ (Equation (27))
       $\mathbf{r}^{(k+1)} \leftarrow \nabla c^{(k+1)}(\mathbf{x}_{dec}) - \nabla c^{(k)}(\mathbf{x}) \mathbb{I}_{(k+1)}^{(k)}$ (Equation (30))
4. $\mathbf{x}^{(k+1)} \leftarrow$ MultigridV($\mathbf{x}_{dec}, \mathbf{r}^{(k+1)}, \mathbf{z}^{(k+1)}, \mathbf{A}^{(k+1)}, k+1$)
5. Perform coarse grid correction
       $\mathbf{x} \leftarrow \mathbf{x} + \mathbb{I}_{(k+1)}^{(k)} (\mathbf{x}^{(k+1)} - \mathbf{x}_{dec})$
       $\mathbf{x} \leftarrow \max\{\mathbf{x}, 0\}$
6. Apply $\nu_2$ ICD iterations with initial condition $\mathbf{x}$ and compute:
       $\mathbf{x} \leftarrow\approx \arg\min_{\mathbf{x}\geq 0} \left\{ ||\mathbf{z} - \mathbf{A}\mathbf{x}||_{\Lambda}^2 + \frac{4^k \hat{\alpha}}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| \frac{x_i^{(k)} - x_j^{(k)}}{2^k} \right|^p - \mathbf{r}\mathbf{x} \right\}$
7. Return($\mathbf{x}$)
}
```

Fig. 6. Pseudo-code specification for the specific V-cycle multigrid inversion algorithm used for the optical tomography problem.

TABLE I

COMPUTATIONAL COMPLEXITY OF THE FIXED GRID ICD/BORN AND THE MULTIGRID INVERSION ALGORITHMS IN TERMS OF NUMBER OF COMPLEX MULTIPLICATIONS PER FULL ITERATION. $M$ = NUMBER OF DETECTORS; $K$ = NUMBER OF SOURCES; $F$ = NUMBER OF ITERATIONS REQUIRED FOR THE LINEAR FORWARD PDE SOLVER; $N$ = NUMBER OF PIXELS; $\nu$ = NUMBER OF THE ICD OPTIMIZATIONS FOR EACH GRID

| | Green's function update | ICD | Total |
|---|---|---|---|
| Fixed Grid | $5(M+K)FN$ | $5MKN$ | $5MKN+5(M+K)FN$ |
| V-cycle | $5(M+K)FN$ | $\frac{20}{3}\nu MKN$ | $\frac{20}{3}\nu MKN+5(M+K)FN$ |
| Full multigrid | $5(M+K)FN$ | $\frac{80}{9}\nu MKN$ | $\frac{80}{9}\nu MKN+5(M+K)FN$ |

TABLE II

ESTIMATES OF THE COMPLEX MULTIPLICATIONS REQUIRED FOR EACH ITERATION OF THE LISTED INVERSION ALGORITHMS

| Parameters | | | | | Fixed grid $\times 10^6$ | V-cycle $\times 10^6$ | Full multigrid $\times 10^6$ |
|---|---|---|---|---|---|---|---|
| $\nu$ | $N$ | $K$ | $M$ | $F$ | | | |
| 2 | $129^2$ | 12 | 12 | 20 | 51 | 71 | 82 |
| 6 | $129^2$ | 12 | 12 | 20 | 51 | 135 | 167 |
| 2 | $129^2$ | 24 | 24 | 20 | 127 | 207 | 250 |
| 6 | $129^2$ | 24 | 24 | 20 | 127 | 463 | 591 |

and forth through coarse and fine resolution in characteristic patterns as shown in Fig. 3(a) and (b).

The V-cycle algorithm is a straightforward generalization of the two grid algorithm. In Figs. 4 and 5 V-cycle algorithm, the fine grid optimization problem is solved by calling the recursive *MultigridV* subroutine starting at the finest grid resolution. Each subroutine call starts by applying $\nu_1$ iterations of a fixed grid optimizer. The *MultigridV* subroutine, then recursively calls itself for the next coarser resolution in order to compute a coarse grid correction to the current fine grid solution. At each coarsening grid resolution, $\nu_1$ iterations are performed using the residual, $\mathbf{r}^{(k)}$, for that level. Finally, the solution is further improved by applying $\nu_2$ iterations of a fixed grid optimizer at each refinement step, using the residuals from the coarsening steps and the

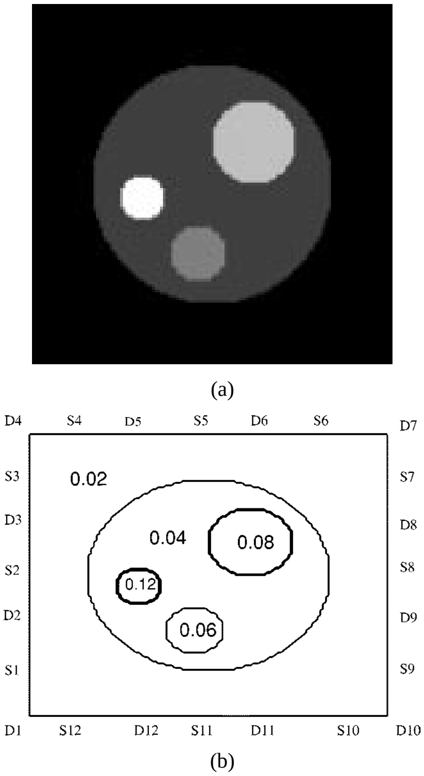


Fig. 7. Two-dimensional 8 cm × 8 cm phantom used for simulation. (a) Gray scale image showing the spatial variation of absorption coefficient and (b) contour plot showing absorption coefficient with units of $\text{cm}^{-1}$.

decimated solution for the coarser grid. This recursive structure causes the algorithm to move from fine to coarse grids, and then back to fine grids, as shown in the "V"' pattern of Fig. 3(a).

The V-cycle algorithm of Fig. 4 includes specific steps to ensure positivity of the result. In particular, the coarse grid correction (Step 5) can potentially result in an $\mathbf{x}$ with negative values. Therefore, the pixel values are limited to a minimum value of 0. We have found this method of enforcing positivity to be very effective in our experiments. However, we note that this step complicates the analysis of the algorithm since it violates the assumptions of our fixed point theorem.

The full multigrid algorithm of Fig. 5 is based on recursive calls of both the full multigrid and V-cycle subroutines. This structure causes the algorithm to initially move to the coarsest grid. The resulting coarsest grid solution is interpolated to the next finer grid and used as the initial condition for the corresponding fixed grid optimization problem, which is then solved by a multigrid V-cycle. This process is repeated, until the final solution is obtained on the finest level.

### C. Multigrid Optimization for the Optical Diffusion Problem

For the optical diffusion tomography problem, we use the ICD optimization method [16], [17], [11], [12] as the fixed grid optimizer at each resolution. The ICD algorithm is a good choice for optimization problems because it has fast convergence at high spatial frequencies [16].

The specific multigrid expressions for the optical diffusion imaging problem are now derived. In all cases, we choose $\mathbb{I}_{(k)}^{(k+1)}$ to be the separable extension of the one-dimensional (1-D) decimation matrix

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \tag{23}$$

and we choose the corresponding interpolation matrix to be

$$\mathbb{I}_{(k+1)}^{(k)} = 4\left(\mathbb{I}_{(k)}^{(k+1)}\right)^T. \tag{24}$$

The finest scale cost function, $c^{(0)}(\mathbf{x}^{(0)})$, is given in (17). At each scale, the cost functional $c^{(k)}(\mathbf{x}^{(k)})$ consists of a quadratic data likelihood term and a nonquadratic prior term. Referring to (19), the quadratic term of (17) can be expressed as

$$\begin{aligned} &\left\|\mathbf{z}^{(k)} - \mathbf{A}^{(k)}\mathbf{x}^{(k)}\right\|_\Lambda^2 \\ &\quad = \left\|\mathbf{z}^{(k)} - \mathbf{A}^{(k)}\left(\hat{\mathbf{x}}^{(k)} + \mathbb{I}_{(k+1)}^{(k)}\left(\mathbf{x}^{(k+1)} - \mathbb{I}_{(k)}^{(k+1)}\hat{\mathbf{x}}^{(k)}\right)\right)\right\|_\Lambda^2 \\ &\quad = \left\|\mathbf{z}^{(k+1)} - \mathbf{A}^{(k+1)}\mathbf{x}^{(k+1)}\right\|_\Lambda^2 \end{aligned} \tag{25}$$

where

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)}\mathbb{I}_{(k+1)}^{(k)} \tag{26}$$

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \mathbf{A}^{(k)}\left(I - \mathbb{I}_{(k+1)}^{(k)}\mathbb{I}_{(k)}^{(k+1)}\right)\hat{\mathbf{x}}^{(k)} \tag{27}$$

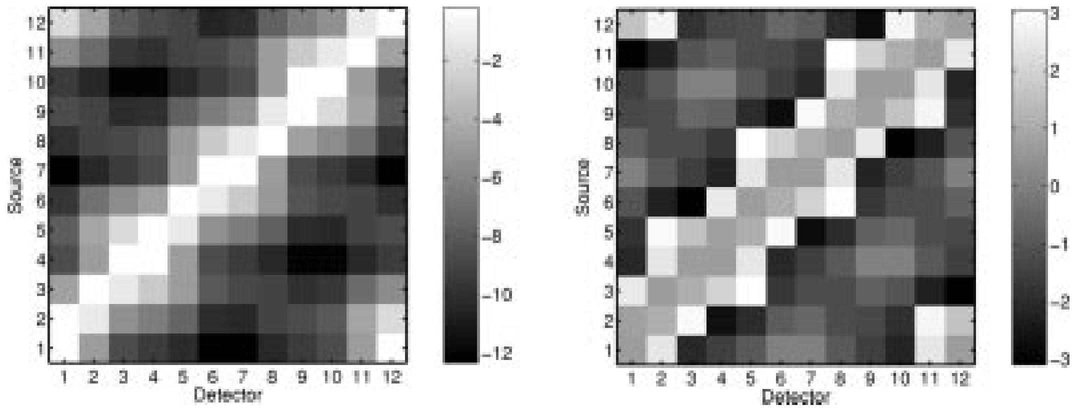
and $I$ denotes the identity matrix.

For the coarse grid prior term, we assume that the derivative of $\mathbf{x}$ is locally smooth [25]. In this case, the corresponding prior term can be represented as

$$\begin{aligned} &\frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|x_i^{(0)} - x_j^{(0)}\right|^p \\ &\quad \simeq \frac{4^k}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|\frac{x_i^{(k)} - x_j^{(k)}}{2^k}\right|^p. \end{aligned} \tag{28}$$

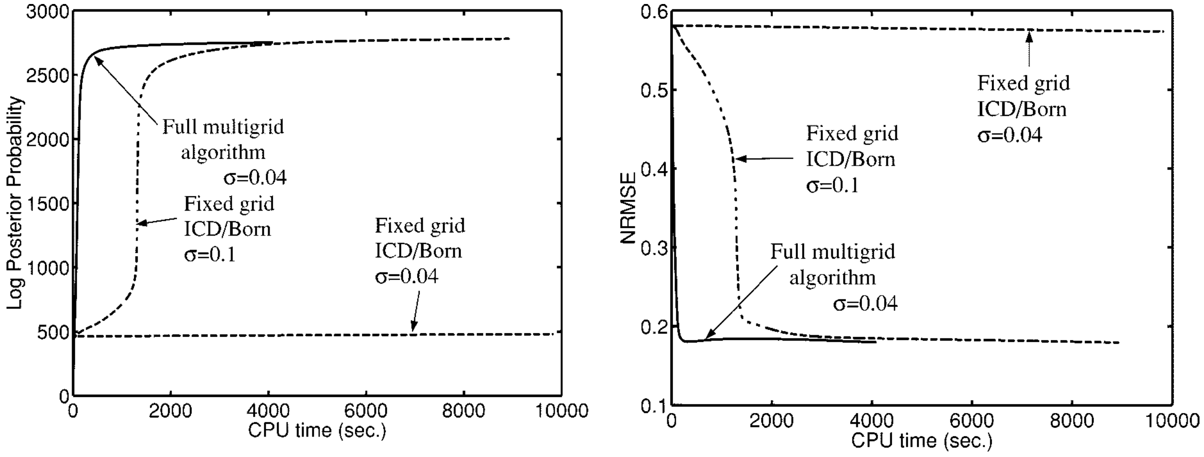
Note that the factor of $4^k$ is chosen to account for the reduced number of terms in the sum, and the smoothness assumption justifies the approximation $x_i^{(0)} - x_j^{(0)} \simeq (x_i^{(k)} - x_j^{(k)})/2^k$. Based on (25) and (28), we then define the coarse grid cost function.

$$\begin{aligned} c^{(k)}\left(\mathbf{x}^{(k)}\right) &= \frac{1}{\hat{\alpha}}\left\|\mathbf{z}^{(k)} - \mathbf{A}^{(k)}\mathbf{x}^{(k)}\right\|_\Lambda^2 \\ &\quad + \frac{4^k}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|\frac{x_i^{(k)} - x_j^{(k)}}{2^k}\right|^p. \end{aligned} \tag{29}$$



(a) (b)

Fig. 8. Gray-scale view of (a) log magnitude and (b) phase of measurements for each source and detector pair. The numbers on the axis denote the indices for the sources and detectors.



(a) (b)

Fig. 9. (a) Log posterior probability and (b) NRMSE as a function of CPU time.

Referring to (22) and (30), the $i$th component of $\mathbf{r}^{(k)}$ is then given by

$$\begin{aligned}\left[\mathbf{r}^{(k+1)}\right]_i &= \frac{2^k}{\sigma^p}\left(2\sum_{j\in\mathcal{N}_i} b_{i-j}\left|\frac{x_i^{(k+1)}-x_j^{(k+1)}}{2^{k+1}}\right|^{p-1}\right.\\ &\quad\cdot\mathrm{sgn}\left(x_i^{(k+1)}-x_j^{(k+1)}\right)-\sum_l\left[\mathbb{I}_{(k+1)}^{(k)}\right]_{l,i}\\ &\quad\left.\cdot\sum_{m\in\mathcal{N}_l} b_{l-m}\left|\frac{x_l^{(k)}-x_m^{(k)}}{2^k}\right|^{p-1}\mathrm{sgn}\left(x_l^{(k)}-x_m^{(k)}\right)\right)\end{aligned} \tag{30}$$

where

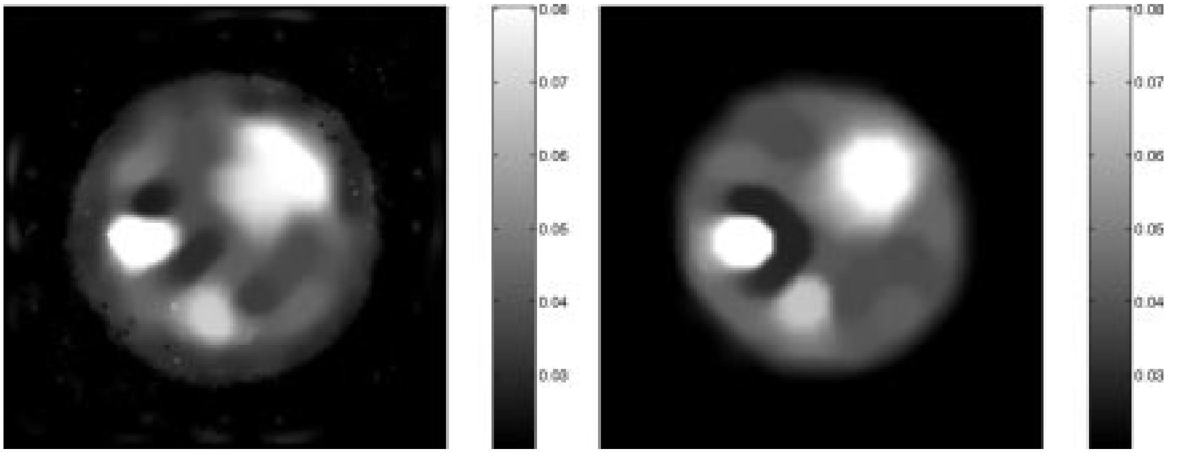
$x^{(k+1)}$ $=\mathbb{I}_{(k+1)}^{(k)}\mathbf{x}^{(k)}$,
$\mathcal{N}_i$ denotes the neighborhood of the $i$th pixel;
$[\mathbb{I}_{(k+1)}^{(k)}]_{i,j}$ is the $(i,j)$th component of the interpolation operator.

Fig. 6 shows the pseudo-code for the multigrid V-cycle subroutine that results for the optical diffusion tomography problem using the cost functions given above.

## IV. COMPLEXITY ANALYSIS

To compare the relative computational costs of the multigrid inversion algorithms with that of the fixed grid ICD/Born algorithm [12], we determine the number of complex multiplications required for one iteration of the V-cycle or the full multigrid inversion algorithm.

Let us assume that the grid resolutions range from $k=0$ to $L-1$, and that the unknown absorption images at grid resolution $k$ are approximately of size $N/4^k$. For simplicity, we neglect the computational cost required for decimation and interpolation of the absorption images. Therefore, the main computational cost is assumed to come for the computation of the Fréchet derivative and the multigrid optimization by the ICD algorithm. We note that the computation of the Fréchet derivative is dominated by solution of the PDE required in evaluating the Green's function, as described in Appendix A.



(a) (b)

Fig. 10. Reconstructions by (a) the fixed resolution ICD/Born algorithm and (b) the full multigrid inversion algorithm.

The V-cycle algorithm performs a total of $\nu = \nu_1 + \nu_2$ fixed grid ICD optimization passes at each grid resolution, where $\nu_1$ and $\nu_2$ are defined in Fig. 4. The number of complex multiplications required for one iteration of ICD is $5MK \times (image\ size)$ [12]. This means that the total computation due to ICD iterations in a single iteration of the V-cycle algorithm, is given by $\sum_{k=0}^{L-1} 5MKN/4^k \leq (20/3)\nu MKN$. In addition, the Fréchet derivative must be computed at the beginning of each multigrid iteration. This adds $5(M+K)FN$ complex multiplications for computation of the Fréchet derivative, where $F$ is the number of iterations chosen for the PDE solver used in the computation of the Green's function [12]. A larger value of $F$ increases the accuracy of the computed Fréchet derivative. We found that $F = 20$ was sufficient for the problems we have studied. The total per iteration computational complexity of the V-cycle algorithm is then listed in Table I as $(20/3)\nu MKN + 5(M+K)FN$.

The full multigrid algorithm performs $\nu k$ ICD iterations at grid resolution $k$. Therefore, the total computation of the ICD iterations at resolution $k$ is $5MKNk/4^k$, and the total ICD computation is therefore bounded by $\sum_{k=0}^{L-1} 5MKNk/4^k \leq (80/9)\nu MKN$. Adding the computation of the Fréchet derivative results in the final expression of $(80/9)\nu MKN + 5(M+K)FN$ listed in Table I.

Table II lists the estimated number of complex multiplications required for each iteration of the fixed grid, V-cycle and the full multigrid inversion algorithms, using typical values of parameters. The point to notice here is that although the number of operations per iteration is larger for the multigrid algorithms than for the fixed grid algorithm, it is not dramatically so. We will see later that the number of iterations required for the multigrid algorithms is substantially less than is required using the fixed grid algorithm, so that overall there is a dramatic decrease in the computation required for the multigrid algorithms.

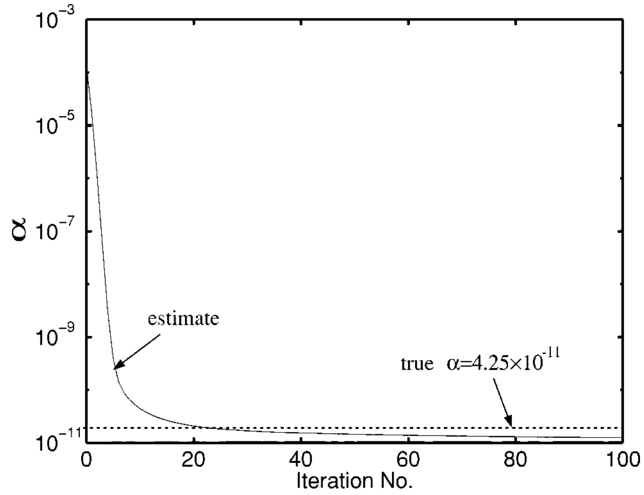


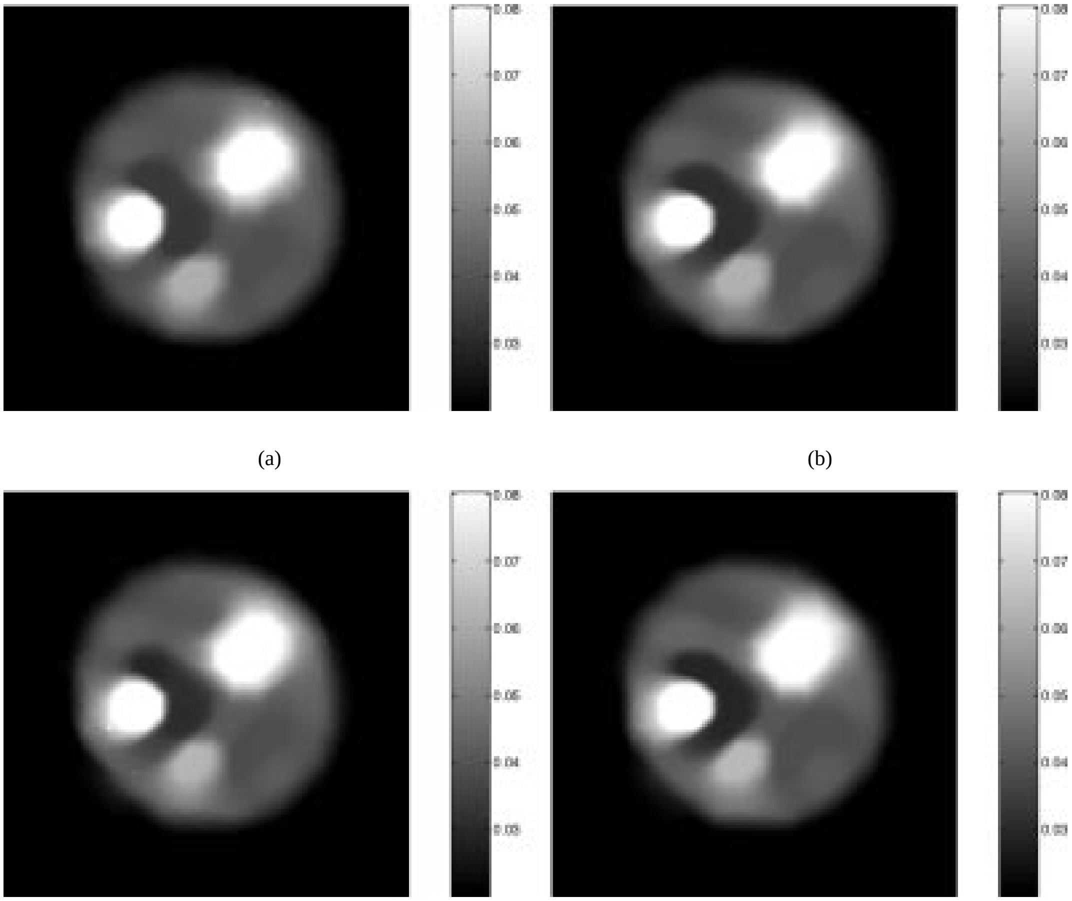
Fig. 11. Estimation of $\alpha$ by (14) as a function of number of iterations of full multigrid algorithm. Note the bias of the estimate from the true value of $\alpha$.

## V. NUMERICAL RESULTS

Simulation results are presented here to assess the performance of the new algorithms. Fig. 7 shows an $8 \times 8$ cm phantom used for one of the numerical experiments. The phantom is discretized on a $129 \times 129$ grid and the absorption coefficient at each grid point, including the background, is considered unknown. The unknown background absorption coefficient is 0.02 cm$^{-1}$, and $\mu_s'$ is assumed uniform throughout $\Omega$ with a value of 10.0 cm$^{-1}$. The values of the absorption coefficient for each sub-domain are given in Fig. 7(b). Fig. 7(b) also shows the locations of the 12 sources and 12 detectors used in the simulations. The modulation frequency is 200 MHz. The synthetic scattering datum for the $k$th source and the $m$th detector pair is generated by adding random noise with a complex Gaussian distribution and with noise variance of $\alpha|\phi_k(d_m)|$ resulting in SNR of [12]

$$\mathrm{SNR}_{mk} = 10\log_{10}\left(\frac{1}{\alpha}\,|\phi_k(d_m)|\right) \tag{31}$$

where $\alpha$ is the noise parameter in (6). This procedure is independently performed for every source and detector pair. The value



(a) (b)

(c) (d)

Fig. 12. V-cycle inversion results with (a) $\nu = 2$, and (b) $\nu = 6$ after 20 iterations; and full multigrid inversion results with (c) $\nu = 2$, and (d) $\nu = 6$ after 20 iterations. CPU times were (a) 352 s, (b) 436 s, (c) 663 s, and (d) 855 s. All the reconstructions are similar and quite accurate.

of $\alpha$ is chosen so that the lowest amplitude measurement has an SNR of 10 dB.

Fig. 8 shows the magnitude and the phase of the simulated data measurements used in the results of Figs. 9–11. Here, the abscissa indexes the detector location, and the ordinate indexes the source location. The gray levels of Fig. 8(a) and (b) are proportional to the log magnitude and phase (from $-\pi$ to $\pi$), respectively. Note that the magnitude peaks at the positions corresponding to nearby source/detector pairs, and attenuates as the distance between source and detectors becomes larger.

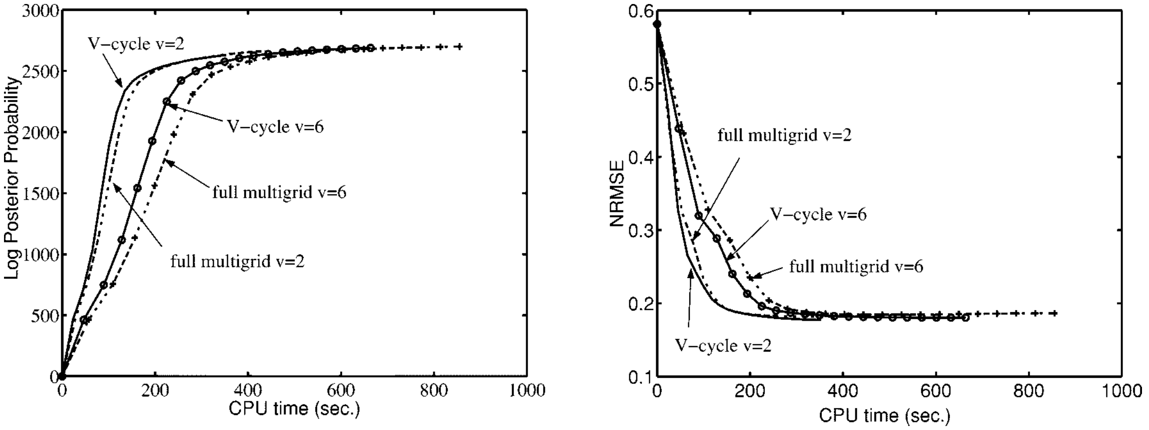
For inversion, we chose an eight-point neighborhood model for the GGMRF prior model with normalized weights $\{b_{i-j}\}$ summing to 1 for each $i$, with $b_{i-j} = (2\sqrt{2}+4)^{-1}$ for nearest neighbors and $b_{i-j} = (4\sqrt{2}+4)^{-1}$ for diagonal neighbors. We used a fixed value of 1.1 for $p$ in all the reconstructions, which has been found suitable for the class of problems considered here [12]. For each iteration of the ICD algorithm, we scanned through the points in a new randomized order. Four grid levels were used in the multigrid algorithm, obtained by decimating the finest grid image with a size of $129 \times 129$ pixels until we obtain a grid of $17 \times 17$ pixels. All reconstructions were initialized with a constant absorption coefficient of 0.02 cm$^{-1}$, corresponding to the background level.

We use the normalized root-mean-square error (NRMSE) of the reconstructed absorption image as a measure of the quality of the reconstructions. The NRMSE is defined as

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^{N}\{\hat{\mu}_a(r_i) - \mu_a(r_i)\}^2}{\sum_{i=1}^{N}\{\mu_a(r_i)\}^2}} \tag{32}$$

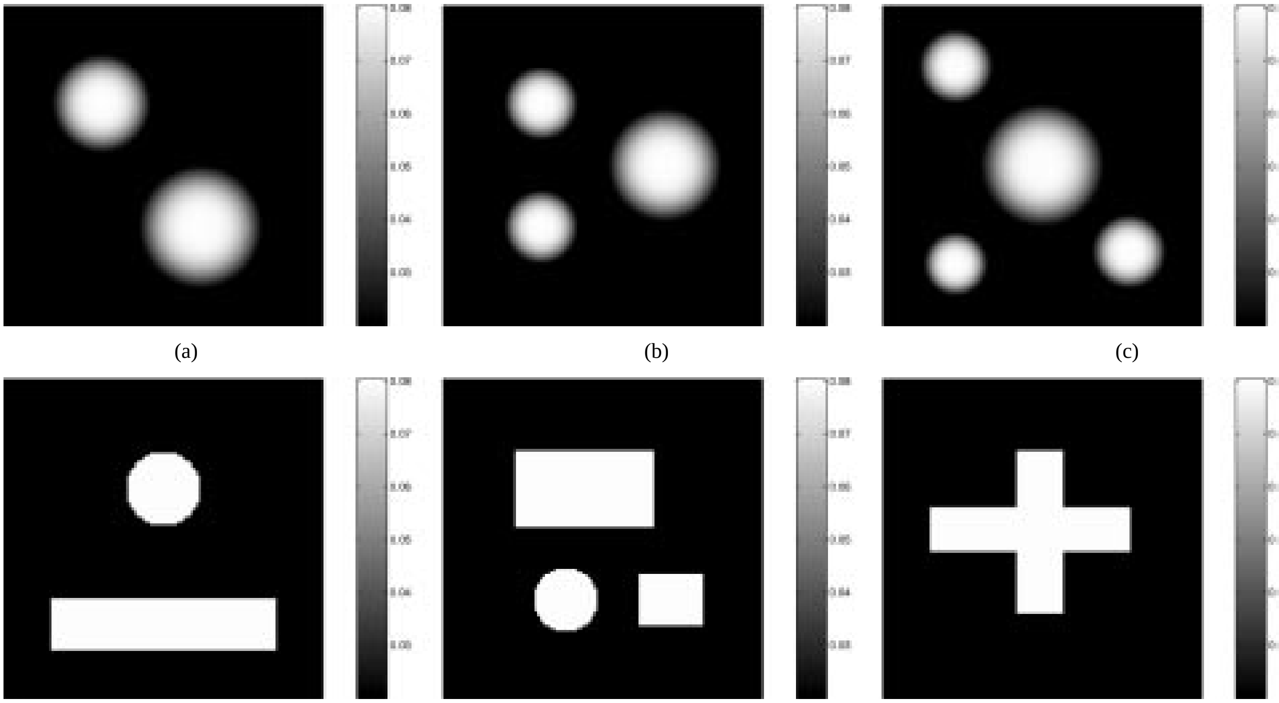
where $\hat{\mu}_a(r_i)$ is the reconstructed value of the absorption coefficient at mesh location $r_i$ and $\mu_a(r_i)$ is the correct value, and the NRMSE is computed at the finest resolution. Furthermore, the log posterior probability (13) is used as a measure of convergence of the algorithms.

Fig. 9 shows the 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. Several hyper-parameter values were tested for the full multigrid solution and it was found that, of these, $\sigma = 0.04\ \text{cm}^{-1}$ qualitatively gave the best result. While the quality of the image should not be a function of the optimization procedure, the convergence of the fixed grid ICD algorithm with $\sigma = 0.04\ \text{cm}^{-1}$



(a) (b)

Fig. 13. (a) Log posterior probability and (b) NRMSE as a function of CPU time. Increasing $\nu$ increases both the NRMSE and the CPU time.



(a) (b) (c)

(d) (e) (f)

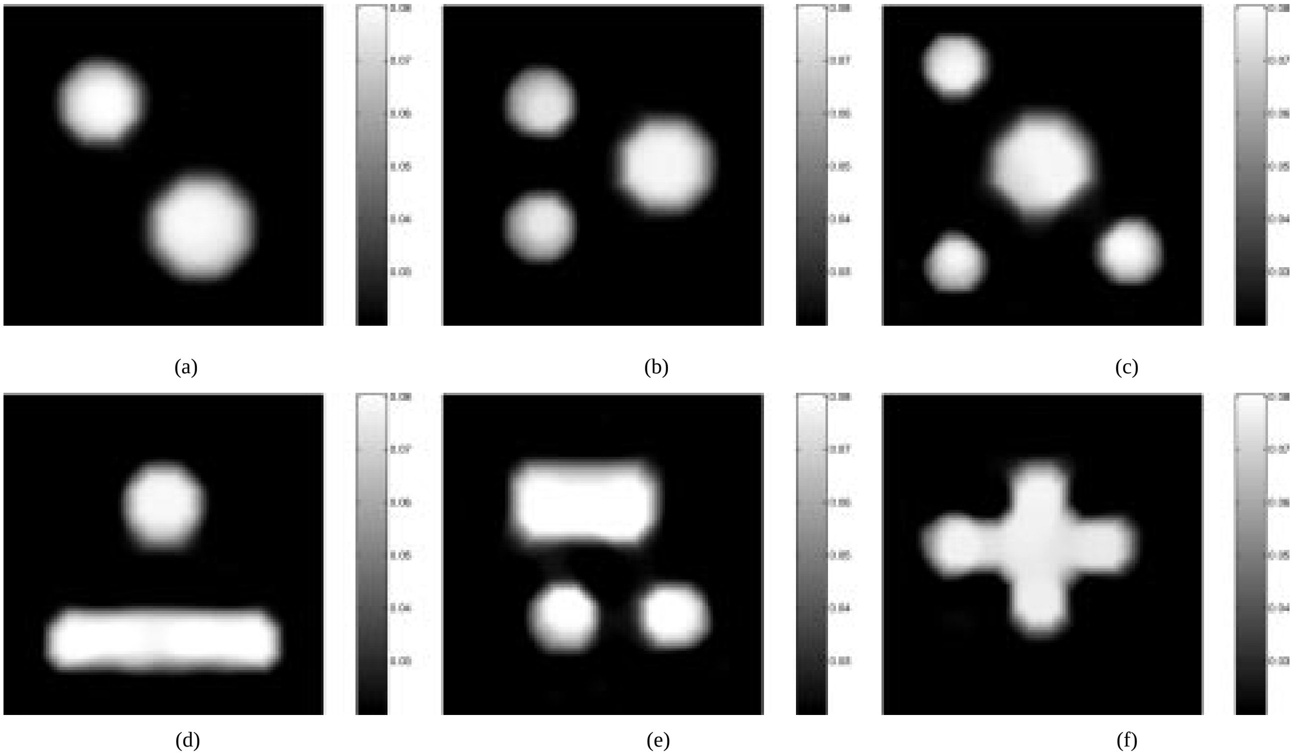
Fig. 14. Variety of absorption image phantoms.

was so poor that after 1000 iterations it still had not achieved a solution close to the optimum achieved by the multigrid algorithm. Therefore, to improve the convergence speed of the fixed grid algorithm, we used a larger value of the hyper-parameter, choosing $\sigma = 0.1\ \mathrm{cm}^{-1}$. This value of $\sigma$ allowed ICD to converge but produced a reconstruction of somewhat lower quality.

Note in Fig. 9 the significant computational savings of the multigrid algorithm over the fixed grid ICD/Born. With the same hyper-parameter value ($\sigma = 0.04\ \mathrm{cm}^{-1}$), the full multigrid solution converges dramatically faster. Even when a larger hyper-parameter ($\sigma = 0.1\ \mathrm{cm}^{-1}$) is used in the fixed grid solution, the multigrid approach is still about 20 times faster. Fig. 10 shows the reconstructions produced by the fixed grid ICD/Born after 1000 iterations (8923 sec of CPU time on a Sun Ultra Sparc 30 machine), and by the full multigrid algorithm after 200 iterations with $\nu = 2$ (4115 s of CPU time). It is evident that the full multigrid algorithm produces a more accurate reconstruction of the phantom. Note that the per-iteration CPU times are 8.93 seconds for the fixed grid algorithm and 20.51 seconds for the multigrid algorithm. This is consistent with the per-iteration complexity listed in Table II. The faster convergence of the multigrid algorithm is due to the substantially fewer iterations required.

Fig. 11 shows the convergence of the $\alpha$ estimation by (14) with respect to the number of full multigrid iterations. The estimation of $\alpha$ also converges rapidly to a constant nonzero value. However, there is a bias in the estimate from the true value of $\alpha$ which was used to generate the synthetic noisy measurement by (32). This is because the joint estimation of $\alpha$ and $\mathbf{x}$ by (14) and (15) produces a biased estimate [42].

Reconstructions using the different recursion patterns of the multigrid inversion algorithms, as well as for different values of



(a) (b) (c)

(d) (e) (f)

Fig. 15. Reconstructions of the phantoms shown in Fig. 14.

$\nu = \nu_1 + \nu_2$ (the total number of ICD optimization passes for each grid), are shown in Fig. 12. Fig. 12(a) and (b) show reconstructions using the V-cycle inversion algorithm after 20 iterations with $\nu = 2$ and $\nu = 6$, respectively, and Fig. 12(c) and (d) show the reconstructions using the full multigrid inversion algorithm after 20 iterations with $\nu = 2$ and $\nu = 6$, respectively. All the reconstructions are similar and quite accurate. The log posterior probability and the NRMSE versus CPU time are shown in Fig. 13. We found that V-cycle or full multigrid with $\nu = 2$ gave slightly better results.

Fig. 15 shows reconstructions for a variety of absorption cross sections (with the true images shown in Fig. 14). In all cases $\mu_s'(r)$ is known and fixed at 10.0 $\text{cm}^{-1}$, the peak values of the absorption coefficient of the inhomogeneities are 0.08 $\text{cm}^{-1}$, and the unknown background is $\mu_a(r) = 0.02\ \text{cm}^{-1}$. The reconstructions are shown for 10 iterations of the full multigrid inversion algorithm with $\nu = 2$, using $p = 1.1$ and $\sigma = 0.02\ \text{cm}^{-1}$. The NRMSE and CPU time after 10 iterations of the full multigrid inversion algorithm given in Table III. The reconstructions are accurate quantitatively and qualitatively, and have a small computational burden (approximately 200 seconds). Note that the NRMSE is higher for Fig. 15(d)–(f). This is because the original images, Fig. 14(d)–(f), have abrupt edges while the original images in Fig. 14(a)–(c) have smoother changes.

## VI. CONCLUSION

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 ill-posed problem, the resulting optimization problem is very computationally expensive.

TABLE III
NRMSE AND CPU TIME FOR THE EXAMPLES OF FIG. 14 AFTER TEN ITERATIONS 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 |

We have developed a general multigrid optimization technique for solving nonlinear inverse problems. This technique incorporates a coarse grid correction scheme to reduce discretization errors. A Bayesian framework has been used for the optical diffusion imaging problem. The algorithm alternately maximizes the log posterior probability with respect to a noise parameter and the unknown image. In each iteration, the noise parameter and Fréchet derivative (calculated using a Born approximation) are updated at the finest grid level. The multigrid optimization is then applied, updating the image by ICD at each grid level.

Simulation results show that the multigrid algorithms dramatically reduce the computational burden as well as improve the reconstruction quality. This improved performance will be essential for realistic 3-D imaging.

## APPENDIX A
## COMPUTATION OF THE FRÉCHET DERIVATIVE

Computation of the Fréchet derivative, $\mathbf{f}'(\mathbf{x})$ for the forward model $\mathbf{f}(\mathbf{x})$ of equation (6) is described here. The Fréchet derivative is a $P \times N$ complex matrix given by

$$\mathbf{f}'(\hat{\mathbf{x}}) =$$

$$\begin{bmatrix} \frac{\partial \phi_1(d_1, \hat{\mathbf{x}})}{\partial x_1} & \frac{\partial \phi_1(d_1, \hat{\mathbf{x}})}{\partial x_2} & \cdots & \frac{\partial \phi_1(d_1, \hat{\mathbf{x}})}{\partial x_N} \\ \frac{\partial \phi_1(d_2, \hat{\mathbf{x}})}{\partial x_1} & \frac{\partial \phi_1(d_2, \hat{\mathbf{x}})}{\partial x_2} & \cdots & \frac{\partial \phi_1(d_2, \hat{\mathbf{x}})}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_1(d_M, \hat{\mathbf{x}})}{\partial x_1} & \frac{\partial \phi_1(d_M, \hat{\mathbf{x}})}{\partial x_2} & \cdots & \frac{\partial \phi_1(d_M, \hat{\mathbf{x}})}{\partial x_N} \\ \frac{\partial \phi_2(d_1, \hat{\mathbf{x}})}{\partial x_1} & \frac{\partial \phi_2(d_1, \hat{\mathbf{x}})}{\partial x_2} & \cdots & \frac{\partial \phi_2(d_1, \hat{\mathbf{x}})}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_K(d_M, \hat{\mathbf{x}})}{\partial x_1} & \frac{\partial \phi_K(d_M, \hat{\mathbf{x}})}{\partial x_2} & \cdots & \frac{\partial \phi_K(d_M, \hat{\mathbf{x}})}{\partial x_N} \end{bmatrix}. \quad (33)$$

In [12], [40], [43], it is shown that each element of the matrix is approximately given by

$$\frac{\partial \phi_k(d_m, \hat{\mathbf{x}})}{\partial x_i} \simeq g(d_m, r_i, \hat{\mathbf{x}}) \phi_k(r_i, \hat{\mathbf{x}}) \cdot \left\{ -1 + \frac{-\hat{\mu}_a(r_i) + j\omega/c}{\hat{\mu}_a(r_i) + \mu_s'(r_i)} \right\} A \quad (34)$$

where

| | |
|---|---|
| $A$ | is the pixel area; |
| is the $\hat{\mu}_a(r_i)$ | current estimate of unknown absorption coefficient at $r_i$; |
| $g(d_m, r_i, \hat{\mathbf{x}})$ | is the Green's function computed as the solution to |

$$\nabla \cdot (\hat{D}(r) \nabla g(r, r_i, \hat{\mathbf{x}})) + (-\hat{\mu}_a(r) + j\omega/c) g(r, r_i, \hat{\mathbf{x}}) = -\delta(r - r_i) \quad (35)$$

with $\hat{D}(r) = 1/3(\hat{\mu}_a(r) + \mu_s'(r))$. Note that $g(d_m, r_i, \hat{\mathbf{x}})$ is the Green's function evaluated at the receiver location $d_m$. In the actual implementation, reciprocity allows us to reduce the computation in the evaluations of the Green's function by interchanging the source location $r_i$ and the detector location $d_m$ [2].

## APPENDIX B
## TWO-GRID FIXED POINT THEOREM

This appendix shows that for an initial value of $\mathbf{x}^{*(k)}$, the two-grid update does not change the solution from this initial value, i.e., $\mathbf{x}^{*(k)}$ is a fixed point of the two-grid update. It is sufficient to show that the unique global minimum of the cost function

$$c^{(k+1)}(\mathbf{x}) - \mathbf{r}^{(k+1)}\mathbf{x} \quad (36)$$

occurs for the value $\mathbf{x} = \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}^{*(k)}$, because in this case the correction term of (19) produces the result $\mathbf{x}^{*(k)}$. To see that this is true, notice that

$$\begin{aligned} &\nabla \left( c^{(k+1)}(\mathbf{x}) - \mathbf{r}^{(k+1)}\mathbf{x} \right) \Big|_{\mathbf{x} = \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}^{*(k)}} \\ &= \nabla\, c^{(k+1)}(\mathbf{x}) \Big|_{\mathbf{x} = \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}^{*(k)}} - \mathbf{r}^{(k+1)} \\ &= \nabla c^{(k)} \left( \mathbf{x}^{*(k)} \right) \mathbb{I}_{(k+1)}^{(k)} \\ &= 0 \mathbb{I}_{(k+1)}^{(k)} = 0 \end{aligned}$$

where the second equality results from (22) and the third equality results from the assumption that $\mathbf{x}^{*(k)}$ is the global minimum of $c^{(k)}(\mathbf{x})$. Since the functional $c^{(k+1)}(\mathbf{x})$ is assumed strictly convex, $\mathbf{x} = \mathbb{I}_{(k)}^{(k+1)} \mathbf{x}^{*(k)}$ must therefore be its unique global minimum.

## 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] W. C. Chew, *Waves and Fields in Inhomogeneous Media*. New York: Van Nostrand Reinhold, 1990.

[3] H. Jiang, K. D. Paulsen, U. L. Osterberg, B. W. Pogue, and M. S. Patterson, "Optical image reconstruction using frequency-domain data: Simulation and experiment," *J. Opt. Soc. Amer. A*, vol. 13, pp. 253–266, Feb. 1996.

[4] 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. Amer. A*, vol. 14, no. 1, pp. 325–342, Jan. 1997.

[5] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. New York: Winston, 1977.

[6] J. E. Dennis, *Numerical Methods for Unconstrained Optimization and Nonlinear Equation*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[7] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.

[8] K. D. Paulsen and H. Jiang, "Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization," *Appl. Opt.*, vol. 35, no. 19, pp. 3447–3458, July 1996.

[9] S. S. Saquib, K. M. Hanson, and G. S. Cunningham, "Model-based image reconstruction from time-resolved diffusion data," in *Proc. SPIE Conf. Medical Imaging: Image Processing*, vol. 3034, Newport Beach, CA, Feb. 25–28, 1997, pp. 369–380.

[10] S. R. Arridge and M. Schweiger, "A gradient-based optimization scheme for optical tomography," *Opt. Express*, vol. 2, no. 6, pp. 213–226, Mar. 1998.

[11] H. Carfantan, A. Mohammad-Djafari, and J. Idier, "A single site update algorithm for nonlinear diffraction tomography," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Munich, Germany, April 21–24, 1997, pp. 2837–2840.

[12] 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. Amer. A*, vol. 16, no. 10, pp. 2400–2412, Oct. 1999.

[13] A. H. Hielscher, A. D. Klose, and K. M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography," *IEEE Trans. Med. Imag.*, vol. 18, pp. 262–271, Mar. 1999.

[14] R. Giering and T. Kaminski, "Recipes for adjoint code construction," *ACM Trans. Math. Softw.*, vol. 24, no. 4, pp. 437–474, Dec. 1998.

[15] S. R. Arridge, "Optical tomography in medical imaging," *Inv. Probl.*, vol. 15, pp. R41–R93, 1999.

[16] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing*, vol. 41, pp. 534–548, Feb. 1993.

[17] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Processing*, vol. 5, pp. 480–492, Mar. 1996.

[18] A. Brandt, *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics*. Sankt Augustin, Germany: GMD-Studien, 1984.

[19] W. Hackbusch, *Multigrid Methods and Applications*. Berlin, Germany: Springer-Verlag, 1985.

[20] S. McCormick, Ed., *Multigrid Methods*. Philadelphia, PA: SIAM, 1987.

[21] S. R. Fulton, P. E. Ciesielski, and W. H. Schubert, "Multigrid methods for elliptic problems: A review," *J. Atmos. Sci.*, vol. 114, pp. 943–959, May 1986.

[22] D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 129–139, Mar. 1986.

[23] J.-M. Laferte, P. Perez, and F. Heitz, "Global nonlinear multigrid optimization for image analysis tasks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Adelaide, Australia, Apr. 19–22, 1994, pp. 533–536.

[24] S. T. Acton, "Multigrid anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, pp. 280–291, Mar. 1998.

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

[26] S. F. McCormick and J. G. Wade, "Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography," *Inv. Prob.*, vol. 9, pp. 697–713, 1993.

[27] M. Bhatia, W. C. Karl, and A. S. Willsky, "Wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. Med. Imag.*, vol. 15, pp. 92–101, Jan. 1996.

[28] 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. Med. Imag.*, vol. 16, pp. 210–217, Apr. 1997.

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

[30] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. New York: Academic, 1978, vol. 1.

[31] M. S. Patterson, B. Chance, and B. Wilson, "Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties," *Appl. Opt.*, vol. 28, no. 12, pp. 2331–2336, June 1989.

[32] M. S. Patterson, J. D. Moultan, B. C. Wilson, K. W. Berndtand, and J. R. Lakowicz, "Frequency-domain reflectance for the determination of the scattering and absorption properties of tissue," *Appl. Opt.*, vol. 30, no. 31, pp. 4474–4476, 1991.

[33] M. O'Leary, D. Boas, B. Chance, and A. Yodh, "Refraction of diffuse photon density waves," *Phys. Rev. Lett.*, vol. 69, no. 18, pp. 2658–2661, 1992.

[34] S. Jacques, "Time resolved propagation of ultrashort laser pulsed within turbid tissues," *Appl. Opt.*, vol. 28, no. 12, pp. 2223–2229, June 1989.

[35] J. S. Reynolds, S. Yeung, A. Przadka, and K. J. Webb, "Optical diffusion imaging: A comparative numerical and experimental study," *Appl. Opt.*, vol. 35, no. 19, pp. 3671–3679, July 1996.

[36] R. H. Kingston, *Detection of Optical and Infrared Radiation*. New York: Springer-Verlag, 1978.

[37] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

[38] T. J. Connolly and D. J. Wall, "On Fréchet differentiability of some nonlinear operators occurring in inverse problems: An implicit function theorem approach," *Inv. Probl.*, vol. 6, pp. 949–966, 1990.

[39] ——, "On an inverse problem, with boundary measurements, for the steady state diffusion equation," *Inv. Prob.*, vol. 4, no. 4, pp. 995–1012, 1988.

[40] 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. Amer. A*, vol. 16, no. 7, pp. 1814–1826, July 1999.

[41] W. Briggs, *A Multigrid Tutorial*. Philadelphia: Soc. Ind. Appl. Math., 1987.

[42] A. Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Lausanne, Switzerland, Sept. 16–19, 1996, pp. 473–476.

[43] J. C. Ye, K. J. Webb, R. P. Millane, and T. J. Downar, "Optimal parameter updating for optical imaging," in *Proc. IEEE Int. Conf. Image Processing*, Chicago, IL, Oct, 4–7, 1998, pp. 390–393.

**Jong Chul Ye** received the B.Sc. (cum laude) and M.Sc. degrees in 1993 and 1995, respectively, from Seoul National University (SNU), Seoul, Korea, both in control and instrumentation engineering, and the Ph.D. degree in 1999 from Purdue University, West Lafayette, IN, in electrical and computer engineering.

He was a Research Assistant with the Department of Control and Instrumentation Engineering, SNU, from 1993 to 1995, and was a Research Scientist with the Korea Advanced System Research Institute from March 1995 to July 1995. From August 1996 to February 1999, he was a Research Assistant with the School of Electrical Engineering, Purdue University. From February 1999 to January 2001, he was a Postdoctoral Research Fellow with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, working on image reconstruction and target shape estimation for passive radar problems. In 2001, he joined the Image Science Laboratory, Polaroid Corporation, Wayland, MA, where he is a Senior Image Scientist working in color science, image and signal processing for digital imaging products such as digital cameras, scanners, and printers.

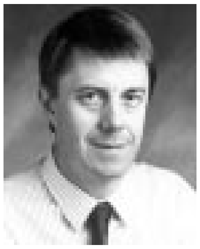
**Charles A. Bouman** (F'00) received the B.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1981 and the M.S. degree from the University of California at Berkeley in 1982. He received the Ph.D. in electrical engineering from Princeton University, Princeton, NJ, in 1989 under the support of an IBM graduate fellowship.

From 1982 to 1985, he was a Full Staff Member with the Lincoln Laboratory, Massachusetts Institute of Technology. In 1989, he joined the faculty of Purdue University, West Lafayette, IN, where he is now a Professor with the School of Electrical and Computer Engineering. His research focuses on the use of statistical image models, multiscale techniques, and fast algorithms in applications such as multiscale image segmentation, fast image search and browsing, and tomographic image reconstruction.

Dr. Bouman is a member of SPIE and IS&T. He has been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, and is currently an Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and is a member of the IEEE Image and Multidimensional Signal Processing Technical Committee. He was a member of the ICIP 1998 organizing committee, and is currently the Vice President for Publications of the IS&T, and a Chair for the SPIE/IS&T conference on Visual Communications and Image Processing (VCIP).

**Kevin J. Webb** (S'81–M'84–SM'98) received the B.Eng. and M.Eng. degrees from the Royal Melbourne Institute of Technology, Melbourne, Australia, in 1978 and 1983, respectively, the M.S.E.E. degree from the University of California, Santa Barbara, in 1981, and the Ph.D. degree from the University of Illinois, Urbana, in 1984.

He is now a Professor with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.



**Rick P. Millane** (S'80–M'82–SM'94) received the B.E. and Ph.D. degrees in electrical engineering from the University of Canterbury, Christchurch, New Zealand, in 1975 and 1981, respectively.

He was a Telecommunications Engineer with the New Zealand Post Office (now Telecom New Zealand) for two years. He joined Purdue University, West Lafayette, IN, in 1982, where he is now a Professor with the Whistler Center for Carbohydrate Research and the Computational Science and Engineering Program. He has held visiting appointments in the Department of Electrical and Electronic Engineering, University of Canterbury (1993, 1999, 2000) and in the Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, CA (1999). His research activities involve inverse problems, phase retrieval, image reconstruction, and diffraction, with applications in biophysical imaging, optical tomography, and ocean acoustics.

Dr. Millane has served as chairman of various scientific meetings, including the OSA Topical Meetings on Signal Recovery and Synthesis (1998 and 2001) and the SPIE Conference on Image Reconstruction from Incomplete Data (2000). He is also active in the American Crystallographic Association.